

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
3 May 2001 (03.05.2001)

PCT

(10) International Publication Number
WO 01/30992 A2

(51) International Patent Classification⁷: C12N 15/00

(74) Agents: HEFNER, M., Daniel et al.; Leydig, Voit & Mayer, Ltd., Two Prudential Plaza, Suite 4900, 180 North Stetson, Chicago, IL 60601-6780 (US).

(21) International Application Number: PCT/US00/29139

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(22) International Filing Date: 20 October 2000 (20.10.2000)

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(25) Filing Language: English

Published:

(26) Publication Language: English

— Without international search report and to be republished upon receipt of that report.

(30) Priority Data:
60/161,092 22 October 1999 (22.10.1999) US
60/227,951 25 August 2000 (25.08.2000) US

(71) Applicant (*for all designated States except US*): UNIVERSITY OF PITTSBURGH OF THE COMMONWEALTH SYSTEM OF HIGHER EDUCATION [US/US]; 200 Gardner Steel Conference Center, Pittsburgh, PA 15260 (US).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(72) Inventor; and

(75) Inventor/Applicant (*for US only*): KOIKE, Chihiro [US/US]; 5628 Hempstead Street, Pittsburgh, PA 15206-1520 (US).



WO 01/30992 A2

(54) Title: α -1-3 GALACTOSYLTRANSFERASE GENE AND PROMOTER

(57) Abstract: The present invention provides a recombinant expression cassette comprising an α -1-3 galactosyltransferase promoter operably linked to a polynucleotide for expression. The invention also provides a recombinant mutating cassette comprising a region of homology to an α -1-3 galactosyltransferase genomic sequence. The cassettes can be employed to express foreign genes or to disrupt the native α -1-3 galactosyltransferase genomic sequence, particularly within an animal. Thus, the invention also provides transgenic animals and methods for their production and use.

α 1-3 GALACTOSYLTRANSFERASE GENE AND PROMOTER

TECHNICAL FIELD OF THE INVENTION

This invention relates to the α 1-3 galactosyltransferase gene, promoters
5 therefor, and the use thereof to create transgenic animals.

BACKGROUND OF THE INVENTION

The current shortage of acceptable organs for transplantation is a major health concern. Because the demand for acceptable organs exceeds the supply,
10 many people die each year while waiting for organs to become available. To help meet this demand, research has been focused on developing alternatives to allogenic transplantation. Thus, for example, dialysis has been available to patients suffering from kidney failure, artificial heart models have been tested, and other mechanical systems have been developed to assist or replace failing organs.
15 Such approaches, however, are quite expensive, and the need for frequent and periodic access to such machines greatly limits the freedom and quality of life of patients undergoing such therapy.

Xenograft transplantation represents a potentially attractive alternative to artificial organs for human transplantation. The potential pool of nonhuman
20 organs is virtually limitless, and a successful xenograft transplantation would not render the patient virtually tethered to machines as is the case with artificial organ technology. Host rejection of such cross-species tissue, however, remains a major concern in this area. Some noted xenotransplants of organs from apes or old-world monkeys (e.g., baboons) into humans have been tolerated for months
25 without rejection. However, such attempts have ultimately failed due to a number of immunological factors. Even with heavy immunosuppression to suppress hyperacute rejection, a low-grade innate immune response, attributable in part to failure of complement regulatory proteins (CRPs) within the graft tissue to control activation of heterologous complement on graft endothelium, ultimately leads to
30 destruction of the transplanted organs (see e.g., Starzl, *Immunol. Rev.*, 141, 213-44 (1994)). In an effort to develop a pool of acceptable organs for
35 xenotransplantation into humans, researchers have engineered animals producing human CRPs, an approach which has been demonstrated to delay, but not eliminate, xenograft destruction in primates (McCurry et al., *Nat. Med.*, 1, 423-27 (1995); Bach et al., *Immunol. Today*, 17, 379-84 (1996)).

In addition to complement-mediated attack, human rejection of discordant xenografts appears to be mediated by a common antigen: the galactose- α (1,3)-

galactose (gal- α -gal) terminal residue of many glycoproteins and glycolipids (Galili et al., *Proc. Nat. Acad. Sci. (USA)*, 84, 1369-73 (1987); Cooper et al., *Immunol. Rev.*, 141, 31-58 (1994); Galili et al., *Springer Sem. Immunopathol.*, 15, 155-171 (1993); Sandrin et al., *Transplant Rev.*, 8, 134 (1994)). This antigen is 5 chemically related to the human A, B, and O blood antigens, and it is present on many parasites and infectious agents, such as bacteria and viruses. Most mammalian tissue also contains this antigen, with the notable exception of old world monkeys and apes (including humans) (see Joziasse et al., *J. Biol. Chem.*, 264, 14290-97 (1989) and references cited therein)). The antigen is highly 10 immunogenic in humans, and many individuals show significant levels of circulating IgG with specificity for gal- α -gal carbohydrate determinants (see, e.g., Galili et al., *J. Exp. Med.*, 162, 573-82 (1985), Galili et al., *Proc. Nat. Acad. Sci. (USA)*, 84, 1369-73 (1987)). Thus, in hopes of better understanding barriers to 15 xenotransplantation, recent attention has turned to the enzyme mediating the formation of gal- α -gal moieties: α 1-3 galactosyltransferase.

The expression of α 1-3 galactosyltransferase is regulated both developmentally and in a tissue-specific manner. The cDNA for this enzyme has been isolated from many species, including pigs (Hoopes et al., poster presentation at the 1997 Xenotransplantation Conference, Nantes France; Katayama et al., *J. Glycoconj.*, 15(6), 583-99 (1998); Sandrin et al., *Xenotransplantation*, 1, 81-88 (1994), Strahan et al., *Immunogenics*, 41, 101-05 (1995)), mice (Joziasse et al., *J. Biol. Chem.*, 267, 5534-41 (1992)), and cows (Joziasse et al., *J. Biol. Chem.*, 264, 14290-97 (1989). While authors have proposed to eliminate the gene from 20 xenograft donor animals (Sandrin et al. (1994), *supra*; U.S. Patent 5,821,117 (Sandrin et al.)), gene knock-out procedures generally require knowledge of the genomic structure and sequence beyond the cDNA of a given gene. The genomic organization of the mouse α 1-3 galactosyltransferase homologue has been deduced (Joziasse et al., *J. Biol. Chem.*, 267, 5534-41 (1992)), and human 25 homologues are known to be inactive pseudogenes (see Joziasse et al., *J. Biol. Chem.*, 266, 6991-98 (1991); Larsen et al., *J. Biol. Chem.*, 265, 7055-61 (1990)). However, the genomic organization of an α 1-3 galactosyltransferase homologue from a species that could serve as a xenograft donor for human recipients has yet 30 to be deduced, and no promoter for any α 1-3 galactosyltransferase homologue gene is known. As such, there exists a need for methods and reagents for 35 facilitating xenotransplantation between species, particularly between species exhibiting differential expression of the gal- α -gal epitope.

BRIEF SUMMARY OF THE INVENTION

The present invention provides a recombinant expression cassette comprising an α 1-3 galactosyltransferase promoter operably linked to a polynucleotide for expression. The invention also provides a recombinant 5 mutating cassette comprising a region of homology to an α 1-3 galactosyltransferase genomic sequence. The cassettes can be employed to express foreign genes or to disrupt the native α 1-3 galactosyltransferase genomic sequence, particularly within an animal. Thus, the invention also provides transgenic animals and methods for their production and use. These aspects of the 10 invention, as well as additional inventive features, will be apparent from the accompanying drawing, sequence listing, and the following detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A through 1I depict the genomic organization porcine α 1-3 15 galactosyltransferase gene. Figure 1A depicts all introns and exons of the gene, indicating the size of the respective elements. Figures 1B through 1I depict alternatively spiced variants isolated from pig aortic endothelial cells.

Figure 2 depicts the organization of a portion of the porcine α 1-3 galactosyltransferase promoter.

20 Figure 3 depicts the organization of the alternate splicing patterns observed in the expression of the human untranslated α 1,3 galactosyltransferase pseudogene.

DETAILED DESCRIPTION OF THE INVENTION

25 In a first aspect, the present invention provides a recombinant expression cassette in which an α 1-3 galactosyltransferase promoter is operably linked to a polynucleotide for expression. The expression cassette is "recombinant" in that within the inventive cassette, the polynucleotide for expression is other than one encoding α 1-3 galactosyltransferase. The promoter and the polynucleotide are 30 "operably linked" in that an event at the promoter (e.g., binding of cellular transcription factors and other DNA binding proteins) precipitates expression (i.e., transcription) of the polynucleotide. So long as this operable linkage is maintained, the cassette can include elements other than the α 1-3 galactosyltransferase promoter and the polynucleotide for expression. For 35 example, the cassette can contain polyadenylation sequences, repressors, enhancers, splice signals, signals for secretion (see, e.g., U.S. Patent 4,845,046 and European Patent EP-B-319,641), etc. Moreover, the expression cassette can

include more than one polynucleotide operably linked to the α 1-3 galactosyltransferase promoter, (e.g., multiple coding sequences separated by internal ribosome entry sites).

- The α 1-3 galactosyltransferase promoter can be derived from any species normally expressing the gene. Thus, for example, the promoter can be derived from the bovine, porcine, or murine α 1-3 galactosyltransferase genes. Examples of such promoters are set forth at SEQ ID Nos:1-6. However, the α 1-3 galactosyltransferase promoter is not limited to one of these sequences, as it can be an active fragment of one of these sequences or a derivative of one of these sequences having one or more mutations (e.g., point mutations, substitutions, insertions, deletions, etc.). Furthermore, given the instant disclosure, it is within the ordinary skill of the art to assay regions of the α 1-3 galactosyltransferase gene unrelated to SEQ ID NOs:1-6 for promoter activity, and the inventive expression cassette can include any α 1-3 galactosyltransferase promoters so identified.
- Suitable promoters can be readily identified by construction an expression cassette in which the derivative sequence is operably linked to a desired reporter gene (e.g., RNA for detection by Northern hybridization, or DNA encoding CAT, luciferase, green-fluorescent peptide, β -galactosidase, etc.) and introducing the cassette into a suitable environment for transcription and (where appropriate) translation.
- Subsequently, promoter activity is detected by assaying for the presence of the reporter by standards methods (e.g., Northern hybridization, Southern hybridization, enzymatic detection, immunohistochemistry, etc.).

Within the expression cassette, the α 1-3 galactosyltransferase promoter can be operably linked to any desired coding polynucleotide. Generally, where expression of a given gene or factor is desired, the skilled artisan will be in possession of the sequence of the coding polynucleotide. Thus, the polynucleotide can be expressed as a bioactive RNA molecule (e.g., an antisense RNA or a ribozyme). Alternatively, the polynucleotide can encode a protein of interest, and in this embodiment, the polynucleotide can be or comprise cDNA or genomic DNA.

Where the polynucleotide encodes a protein, any desired protein can be so encoded, and it need not be syngenic to the species from which the promoter is derived. Thus, for example, the cassette can be employed in animals to produce proteins facilitating growth or bulking of the animal (e.g., bovine or human growth factor) for conferring resistance to disease or parasites. Other encoded proteins can be enzymes such as sulfo- or glycosyltransferases, (e.g., a fucosyltransferase, a galactosidase, a galactosyltransferase, a, a β -acetylgalactosaminyltransferase, an

N-acetylglucosaminyltransferase, an N-acetylglucosaminyltransferase, a sialyltransferase, etc.). Where the expression cassette is employed to generate tissue or organs for xenotransplantation into an organism lacking gal- α -gal antigens (as described below), preferably the polynucleotide encodes a Type I fucosyltransferase, a Type II fucosyltransferase, an α 2-3 sialyltransferase, or an α 2-6 sialyltransferase from any species, the coding sequences of which are known (see, e.g., Larsen et al., *Proc. Nat. Acad. Sci. (USA)*, 87, 6674-78 (1990); Kelly et al., *J. Biol. Chem.*, 270(9), 4640-49 (1995), *J. Biol. Chem.*, 268(30), 22782-87 (1993), Weinstein et al., *J. Biol. Chem.*, 262(36), 17735-43 (1987)).

The expression cassette can be constructed by conventional methods of molecular biology (e.g., direct cloning by ligation, site specific recombination using recombinases, such as the flp recombinase or the cre-lox recombinase system (reviewed in Kilby et al. *Trends Genet.*, 9, 413-21 (1993)), homologous recombination, and other suitable methods). Typically, the promoter sequence is introduced into a vector 5' (i.e., "upstream") of the coding polynucleotide and any other elements (e.g., ribosome entry sites, polyadenylation sequences, etc.), after which the construct is subcloned and grown in a suitable host organism (e.g., yeast, bacteria, etc.) from which it can be isolated or substantially (and typically completely) purified by standard methods. Thus, the invention provides a vector (preferably an isolated or substantially purified vector) including a recombinant expression cassette as set forth above. Such a vector can be any desired type of vector, such as naked DNA vectors (e.g., oligonucleotides or plasmids); viral vectors (e.g., adeno-associated viral vectors (Berns et al., *Ann. N.Y. Acad. Sci.*, 772, 95-104 (1995)), adenoviral vectors (Bain et al., *Gene Therapy*, 1, S68 (1994)), bacteriophages, baculovirus vectors (see, e.g., Luckow et al., *Bio/Technology*, 6, 47 (1988)), herpesvirus vectors (Fink et al., *Ann. Rev. Neurosci.*, 19, 265-87 (1996)), packaged amplicons (Federoff et al., *Proc. Nat. Acad. Sci. USA*, 89, 1636-40 (1992)), papilloma virus vectors, picornavirus vectors, polyoma virus vectors, retroviral vectors, SV40 viral vectors, vaccinia virus vectors) or other vectors (e.g., a cosmid, a yeast artificial chromosome (YAC), etc.). Of course, the vector can (and typically does) contain elements in addition to the expression cassette that are appropriate to the type of vector (e.g., origins of replication, marker genes, genes conferring resistance to antibiotics, etc.). The insertion of the expression cassette can disrupt one or more of these elements, if desired, or the cassette can be inserted between genetic elements to minimize perturbation of the backbone vector.

Where the vector is a viral vector, preferably it is replication incompetent. Thus, for example, an adenoviral vector preferably has an inactivating mutation in at least the E1A region, and more preferably in region E1 (i.e., E1A and/or E1B) in combination with inactivating mutations in region E2 (i.e., E2A, E2B, or both 5 E2A and E2B), and/or E4 (see, e.g., International Patent Application WO 95/34671). An AAV vector can be deficient in AAV genes encoding proteins associated with DNA or RNA synthesis or processing or steps of viral replication (e.g., capsid formation) (see U.S. Patents 4,797,368, 5,354,768, 5,474,935, 10 5,436,146, and 5,681,731). Where the vector is a retroviral vector, the cis-acting encapsidation sequence (E) essential for virus production in helper cells can be deleted upon reverse transcription in the host cell to prevent subsequent spread of the virus (see, e.g., U.S. Patent 5,714,353). Where the vector is a herpesvirus, inactivation of the ICP4 locus and/or the ICP27 cassette renders the virus 15 replication incompetent in any cell not complementing the proteins (see, e.g., U.S. Patent 5,658,724, see also DeLuca et al., *J. Virol.*, 56, 558-70 (1985); Samaniego et al., *J. Virol.*, 69(9), 5705-15 (1996)).

To use the inventive recombinant expression cassette, it is introduced into a eukaryotic cell in a manner suitable for the cell to express the coding polynucleotide. A vector harboring the recombinant expression cassette is 20 introduced into a eukaryotic cell by any method appropriate for the vector employed, which generally are well-known in the art. Thus, plasmids are transferred by methods such as calcium phosphate precipitation, electroporation, liposome-mediated transfection, microinjection, viral capsid-mediated transfer, polybrene-mediated transfer, protoplast fusion, etc. Viral vectors are best 25 transferred into the cells by infecting them.

Depending on the type of vector, it can exist within the cell as a stable extrachromosomal element (which can even be heritable, see e.g., Gassmann, M. et al., *Proc. Natl. Acad. Sci. (USA)*, 92, 1292 (1995)) or it can integrate into the host cell's chromosomes. Thus, the invention provides a chromosome including a 30 recombinant expression cassette such as described above, as well as a cell including such a cassette (and such a chromosome). The α 1-3 galactosyltransferase promoter of the expression cassette can be native to such a cell or chromosome, or it can be exogenous to the cell or chromosome. Where the promoter is native to the cell or chromosome, preferably the polynucleotide for 35 expression within the cassette (the non-native polynucleotide) displaces the operable linkage between the native polynucleotide encoding α 1-3 galactosyltransferase such that it is no longer operably linked to the native α 1-3

galactosyltransferase promoter. Such displacement can be accomplished where the non-native polynucleotide is cloned between the promoter and the native polynucleotide (i.e., upstream of the native polynucleotide), especially where the non-native polynucleotide contains one or more transcriptional termination signals
5 (preferably in all three putative reading frames). Of course, the non-native polynucleotide also can be introduced into the locus such that it destroys the native exon/intron boundaries and/or introduces inactivating mutations (e.g., deletions, insertions, frame-shifts, etc.) into the native coding sequence.

Preferably, the transgenic cell presents a suitable microenvironment for the
10 coding polynucleotide within the expression cassette to be expressed. In many instances, the transgenic cells can be used to study the tissue specificity, dynamics, and kinetics of the promoter, for example by assaying for the expression of the polynucleotide within the cells. However, as the absence of activity is as useful as the presence of promoter activity in these contexts, any cell can be employed for
15 such purposes; such a cell can be *in vivo* or *in vitro*. Preferably, the cell is derived from a species syngenic to the source of the promoter so that, by virtue of the properties of the α 1-3 galactosyltransferase promoter present within the expression cassette, the polynucleotide within the cassette is expressed within such transgenic tissues, organs, or animals with the same kinetics and tissue specificity
20 as the native α 1-3 galactosyltransferase gene in wild-type animals. Where the cells are *in vivo*, they are typically cells of a mammal (e.g., human cells), and can be any type of cells. Suitable cells for use *in vitro* include yeast, protozoa (e.g., *T. cruzi* epimastigotes), cells derived from any mammalian species (e.g., VERO, CV-1, COS-1, COS-7, CHO-K1, 3T3, NIH/3T3, HeLa, C1271, BS-C-1 MRC-5, etc.),
25 insect cells (e.g., *Drosophila* Snyder cells), or other such cells. In other applications, the cell can be employed to construct transgenic tissues, organs, or animals, as described below, in which case the cell typically is a spermatozoon, ovum, zygote, primordial germ cells, or embryonic stem cell.

In another embodiment, the invention provides a method of mutating a
30 region of a chromosome comprising an α 1-3 galactosyltransferase gene. In accordance with the inventive method, a recombinant mutating cassette comprising a region of homology to the α 1-3 galactosyltransferase gene is recombined with a chromosome which has an α 1-3 galactosyltransferase gene such that homologous recombination occurs between the cassette and the
35 chromosome. As a result of the homologous recombination, a mutation is introduced into the native α 1-3 galactosyltransferase chromosomal gene sequence.

Thus, the final step of the method involves screening for successful recombination.

The inventive method employs a recombinant mutating cassette including at least a first region of homology to an α 1-3 galactosyltransferase genomic sequence, and the invention provides such a cassette. Within such a cassette, this first region of homology is adjacent to either to at least one polynucleotide for insertion or to a second region of homology. The mutating cassette is "recombinant" in that neither the second region of homology nor the polynucleotide for insertion is adjacent to the first α 1-3 galactosyltransferase genomic sequence in its native state (i.e., within a chromosome).

The insertion cassette can include more than one polynucleotide for insertion and/or more than one region of homology to all or a portion of the α 1-3 galactosyltransferase genomic sequence. Indeed, where the cassette includes a region for insertion, preferably it has at least two regions of homology flanking the region for insertion. Where more than one region of homology is present, whether adjacent to each other or flanking a region for insertion, the cassette can be used to replace any span of the target chromosomal genomic sequence that lies between the two homologous chromosomal regions. Where multiple regions of homology are present, they should generally be arrayed in the same 5' to 3' orientation relative to one another.

A region of homology can be homologous to any portion of the genomic sequence of an α 1-3 galactosyltransferase gene or the antisense strand thereof. The region can be homologous to the gene of any desired species, such as those discussed above, and it can be homologous to an intron, an exon, a promoter sequence, or any other desired sequence from the genomic DNA. To this end, regions of homology can be selected from the promoter sequences disclosed in SEQ ID NOS:1-6. Alternatively (or additionally) a region of homology can be selected from a portion of the genomic sequence from an α 1-3 galactosyltransferase homologue. In this light, some of the murine sequences have been published (see, e.g., Joziasse et al., *J. Biol. Chem.*, 267, 5534-41 (1992)), and additional portions are set forth as SEQ ID NOS: 17-25. Portions of the porcine genomic sequence are disclosed herein as SEQ ID NOS: 7-16. Portions of the human α 1,3 galactosyltransferase pseudogene genomic sequences are set forth at SEQ ID NOS: 35-42, and various (untranslated) human cDNA transcripts are set forth as SEQ ID NOS: 27-34, and those from Rhesus monkeys are set forth at SEQ ID NOS: 43-44. These sequences disclosed herein, as well as the published murine sequences, include the intron/exon boundaries from which one of skill in

the art can isolate additional intronic genomic sequences by techniques such as genome walking, 5' RACE, 3' RACE, etc.

A region of homology to the genomic sequence of an α 1-3 galactosyltransferase gene need not be an exact complement to the genomic sequence; however, the region must be sufficiently homologous to the α 1-3 galactosyltransferase gene to permit homologous recombination between the cassette and the genomic DNA *in vivo*. Indeed, in some embodiments (e.g., for introducing point mutations into the genomic sequence), a region of homology preferably contains some mismatched bases. Thus, typically, the region of homology will bear at least about 75 % homology to a portion of the α 1-3 galactosyltransferase gene or its antisense strand (such as at least about 85 % homology to a portion of the α 1-3 galactosyltransferase gene or its antisense strand), and more typically the region of homology will bear at least about 90 % homology to a portion of the α 1-3 galactosyltransferase gene or its antisense strand (such as at least about 95 % or even at least about 97 % homology to a portion of the α 1-3 galactosyltransferase gene or its antisense strand). Any commonly employed method (e.g., BLAST database searching) for calculating percent homology can be used to select a suitable region of homology. Similarly, while the length of the region of homology is not critical, it should be sufficiently long to facilitate homologous recombination between the cassette and the genomic DNA *in vivo*. Thus, typically the region of homology will be at least about 50 nucleotides long (such as at least about 75 or 100 bases long), and more typically it will be at least several hundred bases long (such as at least about 250, 500, or even 750 bases long). Indeed, in many applications, the region of homology preferably is several thousand bases long to maximize the likelihood of homologous recombination *in vivo*. The ideal length of a region of homology depends in part on the number of such regions within the cassette – where one or few regions of homology are present, they should be longer to facilitate recombination between the cassette and the genomic DNA; conversely, where the cassette contains several regions of homology, they can be shorter without reducing the likelihood of recombination events.

Where present within the cassette, a region for insertion can be or comprise any DNA which is desired to be introduced into the genomic sequence of an α 1-3 galactosyltransferase gene. Thus, the region can comprise genetic regulatory elements (e.g., enhancers, promoters, repressors, etc., the sequences of which are known) or consensus binding sites for DNA-binding proteins (e.g., restriction endonucleases, transcription factors, etc.). In many applications, a region for

insertion can comprise a polynucleotide for expression, such as those set forth above, or even expression cassettes. A preferred polynucleotide for insertion is an expression cassette for expressing a positive marker flanked by FRT sites, thus facilitating the identification of chromosomes into which the polynucleotide for insertion has integrated as well as excision of the cassette.

The mutating cassette can be constructed by any desirable molecular techniques, and typically, the mutating cassette will be engineered within a vector, such as those set forth above. Typically, the vector is a gene transfer vector suitable for introducing the cassette into a host cell. In addition to the region(s) of homology and the polynucleotide for insertion elements, the mutating cassette can have other components, such as, for example, an expression cassette, a region of homology to other genes or chromosomal regions, a polyadenylation sequence, etc., and it is preferred that the insertion cassette comprises a cassette for expressing at least one marker gene (which may be or comprise the polynucleotide for insertion). Such a marker can be either positive (conferring a visible phenotype to the cells) or negative (killing cells or rendering non-recombinant cells growth-impaired), and both can be used in conjunction. Examples of such positive and negative selection markers are the neosporin resistance (neo^R) gene, the hydromycin resistance (hyg^R) gene, and a thymidine kinase gene (e.g., HSV tk); other suitable markers are known in the art (see, e.g., Mansour et al., *Nature*, 336, 348-52 (1988); McCarrick et al., *Transgen. Res.*, 2, 183-90 (1993)). A marker gene sequence can be bordered at both ends by FRT DNA elements, and/or with stop codons for each of the three putative reading frames being inserted 3' to the desired DNA sequence. Presence of the FRT elements permits the marker to be deleted from the targeted chromosome, and the stop codons ensure that the $\alpha 1,3$ galactosyltransferase gene remains inactivated following deletion of the selectable marker, if inactivation is the desired result of the use of the mutating cassette. The relative orientations of the positive and negative selectable markers are not critical. However, where a positive marker is employed, it should be located between regions of homology, while any negative marker should be outside the regions of homology, either 5' or 3' to those regions.

In accordance with the inventive method, homologous recombination occurs between the $\alpha 1-3$ galactosyltransferase genomic chromosomal DNA and the region (or regions) of homology in the mutating cassette. Where more than one region of homology is present in the cassette, any portion of the genome lying between the homologous target sequences is replaced by whatever sequence lies between the regions of homology in the cassette. Thus, where the mutating

cassette contains a region for insertion flanked by two regions of homology, it will be introduced into the genomic sequence adjacent to the sites of homology, replacing that portion of the genomic sequence. Of course, where the two flanking regions of homology are normally adjacent to each other in the chromosomal sequence, the region for insertion is introduced into the chromosome without replacing any native sequence. Similarly, where no region for insertion is present within the cassette, that portion of the chromosome lying between the two regions of homology in the cassette is deleted as a result of the recombination events. Where the cassette contains a region of homology that differs slightly from the homologous sequence within the genome, it can be employed to introduce point mutations into the genomic sequence.

While the recombination event can occur *in vitro*, typically such homologous recombination occurs within a host cell between an exogenous vector containing the cassette and a chromosome within the host cell containing an α 1-3 galactosyltransferase genomic sequence. Thus, the present invention provides a cell harboring a mutating cassette, as described above. The vector can be introduced into the host cell by any appropriate method, such as set forth above. Commonly, however, the vector is introduced into small cells (e.g., embryonic stem cells) by electroportation and into large cells (e.g., ova or zygotes) by microinjection. Where microinjection is employed, the vector preferably is injected directly into a nucleus or pronucleus of the cell.

The last step in the method is to screen for successful recombination events. Any assay to detect such events can be employed in the context of the inventive method. In accordance with one such assay, chromosomal DNA is screened by PCR or Southern hybridization. For example, where the mutating cassette is designed to delete a portion of the α 1-3 galactosyltransferase genomic sequence, the absence of signal using a probe or primer directed against the region to be deleted indicates a positive recombination event. Conversely, where the cassette includes a region for insertion, a positive result using a probe or primer directed against the region for insertion is indicative of a positive recombination event. Of course, the chromosomal DNA can be sequenced to confirm the correct insertion/deletion/replacement. Where recombination is directed within cells, the events can be screened by assaying for any markers present in the mutating cassette.

By employing the inventive method, one of skill in the art can use the inventive mutating cassette to introduce targeted deletions, insertions, or replacement mutations into any predefined site within the α 1-3

galactosyltransferase genomic sequence. Any desired amount or portion of the gene can be thus deleted, which can lead to complete inactivation of the gene. For introducing inactivating mutations into the gene, preferably at least one region of homology is selected to recombine with the promoter (to inactivate it) or exons 4-9, which contain the coding sequences. Similarly, the inventive method can introduce functional expression cassettes in place of the α 1-3 galactosyltransferase gene, which can be under the control of the native α 1-3 galactosyltransferase promoter or an exogenous promoter within the cassette (especially where the native α 1-3 galactosyltransferase promoter is destroyed). Thus, the present invention provides a recombinant chromosome containing such a mutation, and a recombinant cell comprising such a chromosome.

As mentioned above, the invention provides recombinant cells and chromosomes comprising a recombinant expression cassette comprising an α 1-3 galactosyltransferase promoter or a mutating cassette, as described above. Indeed, as a result of using these reagents and methods, the invention also provides a cell having a mutant α 1-3 galactosyltransferase genomic sequence, as described above. While any cell having such exogenous genetic sequences is within the scope of the invention, preferably the cells are suitable for constructing a recombinant animal, and are most preferably totipotent cells. Thus, preferred cells are embryonic stem (ES) cells, ova, primordial germ cells (PGCs), and zygotes. ES cells and PGCs are especially preferred because such cells can be obtained and cultured in relatively large numbers relative to ova and zygotes. Using such cells, a transgenic animal having an expression cassette comprising an α 1-3 galactosyltransferase promoter or a disruption in this gene can be constructed by methods known in the art (see e.g., U.S. Patents 5,850,004 (MacMicking et al.), 5,942,435 (Wheeler), 5,523,226 (Wheeler), and 5,175,383; White et al., *Transplant. Int.*, 5, 648-50 (1992); McCurry et al., *Nat. Med.*, 1, 423-427 (1995); Hogan et al., *Manipulating the Mouse Embryo*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1986); Hammer et al., *Nature*, 315, 680 (1985); Murray et al., *Reprod. Fert. Devl.*, 1, 147, (1989); Purselet et al., *Vet. Immunol. Histopath.*, 17, 303 (1987); Rexroad et al., *J. Reprod. Fert.*, 41, (suppl.), 119 (1990); Rexroad et al., *Molec. Reprod. Devl.*, 1, 164 (1989); Simonset et al., *BioTechnology*, 6, 179 (1988); Vizeut et al., *J. Cell. Sci.*, 90, 295 (1988); Wagner, *J. Cell. Biochem.*, 13B (suppl.), 164 (1989); Thomas et al., *Cell*, 51, 503 (1987); Capecchi, *Science*, 244, 1288 (1989); Joyner et al., *Nature*, 338, 153 (1989); Ausubelet et al., *Cur. Prot. Mol. Biol.*, John Wiley & Sons (1987)).

- Where ova and zygotes are employed, after the introduction of the cassette, they can be implanted into surrogate mothers to develop into adult animals.
- Where ES cells or PGCs are employed, after the introduction of the cassette, they typically are further manipulated (e.g., by injection into a blastocyst or morula,
- 5 co-culture with a zona pellucida-disrupted morula, fusion with an enucleated zygote, etc.) such that their mitotic descendants are found in a developing embryo. Such an embryo typically is a chimera composed of normal embryonic cells as well as mitotic descendants of the introduced ES cells or PGCs. Alternatively, the genome of an ES cell or PGC can be incorporated into an embryo by fusing the ES
- 10 cell/PGC with an enucleated zygote to create a non-chimeric embryo in which all nuclei are mitotic descendants of the fused ES cell/PGC nucleus. In any event, to produce a transgenic animal, the embryo or zygote is implanted into a pseudopregnant animal, which, after suitable gestation, gives birth to an animal containing the mutant chromosome containing the cassette in its germ line (if a
- 15 chimera) or possibly all of its cells. Of course, as mentioned above, where the animal is engineered to include a non-mutating expression cassette, it can be inherited as an extrachromosomal plasmid (Gassmann, M. et al., *supra*)). However constructed, the presence of the recombinant allele can be confirmed by performing Northern hybridization or rt-PCR on RNA isolated from the animal in question.
- 20 After birth and sexual maturation, a chimeric animal can be mated to generate a heterozygous animal comprising a disrupted α 1-3 galactosyltransferase gene or recombinant expression cassette (integrated or extrachromosomal) including a α 1-3 galactosyltransferase promoter. Heterozygotes can be crossed to produce a homozygous strain. Such animals having a recombinant expression cassette including an α 1-3 galactosyltransferase promoter, as discussed above, will express the polynucleotide for expression of such cassette within the same tissue types and with the same kinetics as a wild-type animal of the same species and strain expresses the α 1-3 galactosyltransferase gene. Of course, homozygous
- 25 transgenic animals of the present invention having a disruption in the α 1-3 galactosyltransferase gene will produce altered forms of the protein or no functional protein at all. Desirably, the phenotype of such "knock out" animals relative to an animal having a wild type α 1-3 galactosyltransferase gene is a markedly increased time of survival of cells isolated or derived from the
- 30 transgenic animal in the presence of human serum, which can be assessed by any desired method (see, e.g., Osman et al., *Proc. Nat. Acad. Sci. (USA)*, 94, 14677-82 (1997)).

The inventive transgenic animals are useful for any use to which animals can be put, and they can be any desired species (e.g., pigs, cows, mice, cats, dogs, etc.). Transgenic mice in which a reporter gene is operably linked to the α 1-3 galactosyltransferase promoter are valuable reagents for assessing the activity and specificity of the promoter. Transgenic livestock (e.g., pigs, cows, goats, and the like) having an inventive expression cassette in which a growth hormone is expressed under the control of the α 1-3 galactosyltransferase promoter can be matured or bulked better than commonly employed strains. Tissue obtained from a transgenic animal according to the present invention can be implanted into a host according to standard surgical methods, and the invention concerns a method of xenotransplantation from a transgenic animal as described herein. The invention also provides a transgenic organ consisting essentially of transgenic cells engineered as described above (e.g., a lung, a heart, a liver, a pancreas, a stomach, an intestine, a kidney, a cornea, skin, etc.), particularly for use in the method of transplantation. The host can be any animal host, such as a pig, a dog, a cat, a cow, a goat, etc. Of course, the recipient can be a human as well, in which case the source animal preferably is a pig.

Transgenic animals lacking a functional α 1-3 galactosyltransferase gene are attractive sources of organs and tissues for xenotransplantation into primates, especially humans, because the tissues of such animals lack the highly antigenic gal- α -gal epitope. Similarly, transgenic pigs having a recombinant expression cassette in which a coding sequence for Type I fucosyltransferase, a Type II fucosyltransferase (especially α (1,2) fucosyltransferase), an α 2-3 sialyltransferase, or an α 2-6 sialyltransferase is operably linked to the α 1-3 galactosyltransferase promoter also are suitable sources of xenotransplantation tissues, as these encoded enzymes compete for the same substrate as α 1-3 galactosyltransferase, and their presence can reduce (preferably below an antigenic threshold) the gal- α -gal antigens in tissues derived from such animals. Indeed, α (1,2) fucosyltransferase converts this substrate into the universally-tolerated H antigen (i.e., the "O" blood-type antigen) and also blocks the addition of the α 1,3 gal moiety. As such, a gene encoding α (1,2) fucosyltransferase is an especially preferred polynucleotide for expression to be included within the inventive recombinant expression cassette. A preferred source animal for xenotransplantation tissues (and by extension the tissues themselves) preferably contains a disruption in the α 1-3 galactosyltransferase gene as well as having a recombinant expression cassette in which a coding sequence for Type I fucosyltransferase, a Type II fucosyltransferase (especially α (1,2)

fucosyltransferase), an α 2-3 sialyltransferase, or an α 2-6 sialyltransferase is operably linked to the α 1-3 galactosyltransferase promoter. More preferably, the animal contains a disruption in the native promoter of α 1-3 galactosyltransferase and an α (1,2) fucosyltransferase coding sequence under the control of its own 5 promoter. Most preferably, the source animal also expresses exogenous human complement regulatory proteins, as discussed above, to further minimize host resistance of the xenograft tissue.

It will be apparent that a transgenic animal created in accordance with the invention can have the exogenous gene cloned in place of the native α 1,3 10 galactosyltransferase gene (i.e., a "knock-in" approach). Indeed, in many embedment such a "knock-in" approach is preferable, for example to avoid the potential of the development of congenital cataracts in purely "knock-out" animals (e.g., as a result of opportunistic infections of microbes bearing the gal- α -gal motif). Indeed, such an approach can afford a safe alternative to broadband 15 antibiotics in livestock and pets, a current public health concern. In this respect, the invention can be employed to create heartier and healthier livestock and pets.

While one of skill in the art is fully able to practice the instant invention upon reading the foregoing detailed descriptions, in conjunction with the drawing and the sequence listing, the following examples will help elucidate some of its 20 features. In particular, these examples indicate how the genomic structure of the porcine α 1-3 galactosyltransferase gene is elucidated, and how the identity and activity of the α 1-3 Galactosyltransferase promoter is assessed. As these examples are presented for purely illustrative purposes, they should not be used to construe the scope of the invention in a limited manner, but rather should be seen 25 as expanding upon the foregoing description of the invention as a whole.

Many experiments described in these examples employed well known techniques and reagents (see, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2d edition, Cold Spring Harbor Press (1989)). Accordingly, in the interest of brevity, the examples do not present the experimental protocols in 30 detail. In the experiments, enzymatic isolation and culture of porcine aortic endothelial cells (PAEC) was performed. PAEC were maintained in Dulbecco's modified essential medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 10,000 units of Heparin (ELKINS-SINN, Inc., Cherry Hill, NJ), 15 mg of endothelium growth supplement (Collaborative Biomedical Product Inc., Bedford, 35 MA), L-glutamine, and penicillin-streptomycin. RNA was obtained from the organs of pigs (Brain, Heart, Spleen, Gut, and Thymus) and PAEC using Trizol reagent

(Gibco Ltd.,). Primers used to clone and identify regions of the porcine, murine, human, and Rhesus monkey genes are set forth at SEQ ID NOS: 45-96.

Example 1

5 This example describes the identification of the 5' untranslated region and genomic structure of the porcine α 1-3 galactosyltransferase gene.

A comparison of published sequences for the α 1-3 galactosyltransferase cDNA (Hoopes et al., *supra*, Katayama et al., *supra*; Sandrin et al., *supra*; and Strahan et al., *supra*) revealed a divergence in the 5' boundary. Some of these 10 cDNA contain putative 5' untranslated sequences that bear a high (> 70 %) homology to murine sequences identified as the second exon, and it was hypothesized that this region is conserved as an exon in the porcine genome as well.

Further 5' sequence was cloned using 5' RACE, and the putative 15 transcription initiation site was probed by S1 protection assay, using standard protocols. Briefly, a plasmid containing the upstream genomic sequence was digested with restriction enzyme, Pml I, and linearized. The DNA was phosphorylated with shrimp alkaline phosphatase, heated to inactivate the enzyme, and then precipitated with ethanol. The linearized plasmid was digested again 20 with Bgl II to yield a probe fragment, which was then end-labeled with γ -³²P-ATP.

The probe was purified using G-25 sephadex, and about 16 μ l was mixed with 20 μ g of total RNA from pig aortic endothelial cells (PAEC), pig brain, and yeast (control), and the aliquots were coprecipitated using NH₄OAc and ethanol. Pellets were resuspended in a standard hybridization buffer, heated to 95 °C for 3- 25 4 minutes, and then incubated at 42 °C overnight.

After incubation, the yeast sample was split into two aliquots, and to each was added a standard S1 nuclease buffer. S1 nuclease was added to one aliquot, while the other did not receive the enzyme. The PAEC and brain samples each received the enzyme and the buffer. All samples were incubated for 30 minutes at 30 37 °C, after which the reactions were stopped by the addition of a standard S1 inactivation buffer. Following the reaction, the samples were then precipitated, resuspended in 5 μ l of a standard gel loading buffer, and resolved using a 6% denaturing polyacrylamide gel.

The data revealed at least 8 separate alternatively spliced transcripts from 35 PAEC, and additional splicing patterns from brain transcripts. Analysis of these sequences revealed three potential upstream exons (1, 1A, and 2), the boundaries of which comply with the AG-GT consensus, and six coding exons (4-9) also were

identified, which agreed with published results. Interestingly, the pig sequence seemingly lacks upstream exon 3 of the mouse 5' untranslated region. The overall organization of the pig genome is depicted in Figure 1. Alternatively spliced forms isolated from PAED are indicated in Figures 1B though 1I. Exon 1A is
5 observed in transcripts isolated from brain tissue.

As mentioned, the transcripts obtained from PAEC and brain revealed several alternative splicing patterns. Using the genomic clone, intronic sequences were identified by "gene walking" using the method and reagents supplied with the UNIVERSAL GENEOMEWALKER™ KIT (Clontech Labs., Inc.). Primers
10 (Seq ID NOs:41-56) were designed to hybridize with the cDNA, and also to the adapter sequence supplied with the Clontech kit. A series of nested PCR reactions was then performed to clone SEQ ID NOs:7-16, which were sequenced. From these results, the intron/exon boundaries were elucidated.

Summing the nucleotides of all identified exons predicts a transcript of about 3.8 kb. This prediction was assessed by Northern analysis. 20 µg of total RNA from PAEC, and pig brain, heart, spleen, gut, and thymus, were respectively separated on formamide agarose gels, and electrotransferred onto nylon membrane. The blots were hybridized with radiolabeled probes ($2.5\text{-}4.0 \times 10^4$ cpm/ml) specific for pig GT exon 1 and exon 9 identified. The blots were exposed to Bio-MAX films (Eastman Kodak Co., Rochester, NY) for 6 days with intensifying screen. The results revealed primary transcripts of between 3.5-3.8 kb, in accordance with the predicted size and the published size for the bovine transcript.
20

25 Example 2

This example describes the identification of the 5' untranslated region and organization of the murine α 1-3 galactosyltransferase gene.

To identify the 5' and 3' ends of α 1,3GT gene transcripts, 5'- and 3'-RACE procedures were performed using the Marathon cDNA Amplification Kit
30 (Clontech) with the spleen poly A⁺ RNA of Balb/C adult male as template. To identify exon-intron boundaries or 5'- and 3'-flanking region of the transcripts, Murine GenomeWalker libraries were constructed using the Universal GenomeWalker Library Kit (Clontech) with Balb/C genomic DNA.

The results of these experiments revealed several genomic sequences,
35 which are set forth at SEQ ID NOs: 17-25. The deduced 5' untranslated nucleotide sequences are longer by 56 bp than previously reported (Joziasse et al., *J. Biol. Chem.*, 267, 5534-41 (1992). The relative intensity of Luciferase activity

by the pGL3/1280 construct was 15-fold higher than that of pGL3-Basic. The 3'-RACE revealed an extended 3'-UTR sequence 30 bp more than previously reported (*Id.*), but no other 3' UTR exon usage. The overall length of the transcript was 2586 bp, 89 bp longer than previously reported (*Id.*).

5 An overall comparison of 5'-UTR of cDNA sequences of the porcine (747 bp) and murine (492 bp) α 1,3GT gene indicates that the homology is observed only in the region of exon 2 (71.7%). Exon 3 observed in mice is not observed in the pig. Murine exon 1 shows no homology with porcine exon 1.

10 **Example 3**

This example describes the identification of the organization of the human and Rhesus monkey α 1-3 galactosyltransferase untranslated pseudogene.

Working from published partial sequence of the human α 1,3 GT ninth exon, primers were designed to identify the start and end of the gene by 5'-RACE, 15 3'RACE and rtPCR, as described above. Several alternate transcripts were identified, and these are set forth as SEQ ID NOs:27-34. The sequences were compared to those of other species employing a formula based on the consensus motif of the splicing acceptor junction: total number of pyrimidines plus 1 (for a branched A) among forty nucleotides per junction. Intron exon boundaries were 20 confirmed as discussed above (see SEQ ID NOs: 35-42). The organization of the alternative splicing patterns observed is indicated in Figure 3.

Using similar techniques, primers were designed based on a partial published sequence (Genbank Accession No. M73306) having homology to exon 9. Initially, 3'RACE showed only poly-A tails, evidence that transcripts exist. 5'- 25 RACE results revealed sequences of high homology to those α 1,3 sequences previously identified (e.g., porcine, bovine and murine), consistent with the identity of the sequence as the Rhesus pseudogene. The sequence of the Rhesus monkey transcripts are set forth at SEQ ID NOs: 43 and 44.

30 **Example 4**

This example describes the identification of the porcine, murine, and bovine α 1-3 galactosyltransferase promoters.

Using PCR and restriction digestions, various sized fragments between nucleotides 1981 and 2992 of SEQ ID NO:7 (porcine) and between nucleotides 35 375 and 1325 (murine) were generated. The fragments were cloned into a plasmid such that they were operably linked to a luciferase coding sequence. PAEC were then transfected with these constructs and probed for luciferase activity, along

with a positive and a negative (no promoter) control. All fragments exhibited significantly greater promoter activity over the negative control (between about 15% and 90 % relative light units, as compared to the positive control, the negative control exhibiting no luciferase activity). These results indicate that the
5 regions are promoters and that the 5'-RACE results discussed in Examples 1 and 2 most likely represent the potential transcription initiation site (TIS). Moreover, sequence analysis of these regions reveals the presence of at least 8 SP1 or GC boxes within it and potentially seven AP-2 consensus binding motifs see also Figure 2). This suggests that the gene may contain alternative start sites, and that
10 sequences within exon 1 may also contain promoter activity. Other sequences from which α 1,3 GT promoters can be derived are set forth as SEQ ID NOs:1-6.

All of the references cited herein, including patents, patent applications, and publications, are hereby incorporated in their entireties by reference.

15 While this invention has been described with an emphasis upon preferred embodiments and illustrative examples, it will be obvious to those of ordinary skill in the art that variations of the preferred embodiments may be used and that it is intended that the invention may be practiced otherwise than as specifically described herein. Accordingly, this invention includes all modifications
20 encompassed within the spirit and scope of the invention as defined by the following claims.

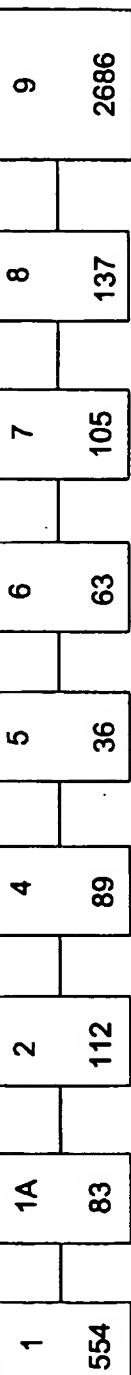
WHAT IS CLAIMED IS:

1. A recombinant expression cassette comprising an α 1-3 galactosyltransferase promoter operably linked to a polynucleotide for expression, other than a polynucleotide encoding α 1-3 galactosyltransferase.
- 5 2. The recombinant expression cassette of claim 1, wherein said promoter is derived from the bovine, porcine, or murine α 1-3 galactosyltransferase genes.
3. The recombinant expression cassette of claim 1 or 2, wherein said promoter comprises any of SEQ ID Nos:1-6.
4. The recombinant expression cassette of any of claims 1-3, wherein said 10 promoter comprises an active derivative of any of SEQ ID Nos:1-6.
5. The recombinant expression cassette of any of claims 1-4, wherein said polynucleotide for expression encodes an antisense RNA molecule or a ribozyme.
6. The recombinant expression cassette of any of claims 1-5, wherein said polynucleotide for expression encodes a protein.
- 15 7. The recombinant expression cassette of claim 6, wherein said protein is a fucosyltransferase, a galactosyltransferase, a β -acetylgalactosaminyltransferase, an N-acetylglycosaminyltransferase, an N-acetylglucosaminyltransferase, a sialyltransferase, or a sulfotransferase.
8. The recombinant expression cassette of claim 6, wherein said protein is a 20 Type I fucosyltransferase, a Type II fucosyltransferase, an α 2-3 sialyltransferase, or an α 2-6 sialyltransferase.
9. The recombinant expression cassette of any of claims 6-8, wherein said polynucleotide for expression is heterogenic to said promoter.
10. The recombinant expression cassette of claim 9, wherein said 25 polynucleotide for expression is human and wherein said promoter is porcine.
11. The recombinant expression cassette of any of claims 6-8, wherein said polynucleotide for expression is a cDNA.
12. The recombinant expression cassette of any of claims 6-8, wherein said polynucleotide for expression is genomic DNA.
- 30 13. A recombinant mutating cassette comprising a first region of homology to an α 1-3 galactosyltransferase genomic sequence adjacent to either a second region of homology to said α 1-3 galactosyltransferase genomic sequence or a polynucleotide for insertion.
14. The recombinant mutating cassette of claim 13, comprising first and 35 second regions of homology to an α 1-3 galactosyltransferase genomic sequence flanking a polynucleotide for insertion.

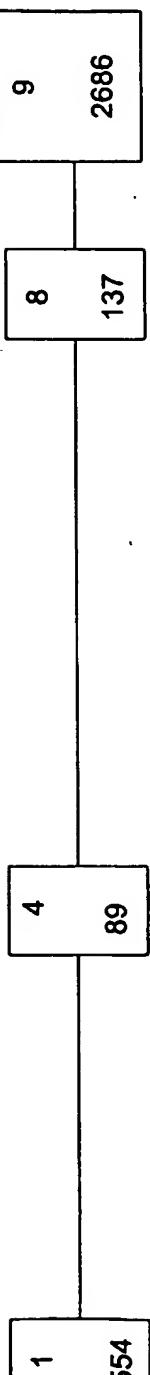
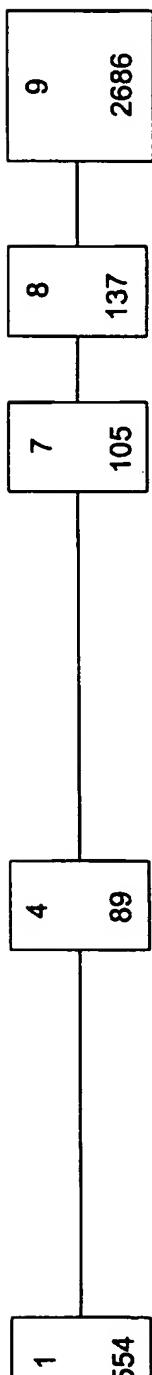
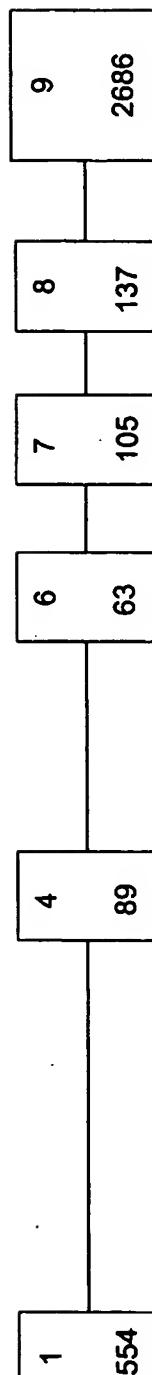
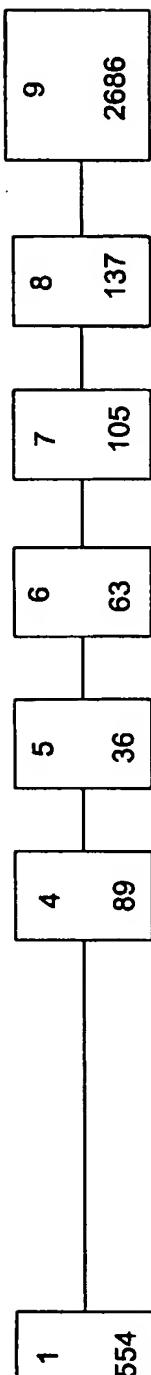
15. The recombinant mutating cassette of claim 13 or 14, wherein a region of homology is homologous to an exon, an intron, or a promoter of said α 1-3 galactosyltransferase genomic sequence.
16. The recombinant mutating cassette of any of claims 13-15, wherein a region of homology is homologous to all or a portion of any one of SEQ ID NOS: 1-42.
17. The recombinant mutating cassette of any of claims 13-16, wherein said polynucleotide for insertion comprises an expression cassette.
18. The recombinant mutating cassette of claim 17, wherein said expression cassette encodes a marker.
19. A vector comprising the recombinant cassette of any of claims 1-18.
20. The vector of claim 19, which is an oligonucleotide, a plasmid, a cosmid, or a virus.
21. A transgenic cell harboring the vector of claim 19 or 20.
22. A chromosome comprising the recombinant expression cassette of any of claims 1-18.
23. A transgenic cell harboring the chromosome of claim 22.
24. The transgenic cell of claim 23, wherein said α 1-3 galactosyltransferase promoter is native to said cell.
25. The transgenic cell of claim 23 or 24, wherein said polynucleotide for expression displaces a native polynucleotide encoding α 1-3 galactosyltransferase.
26. The transgenic cell of claim 23 or 24, wherein said polynucleotide for expression is cloned between said promoter and a native polynucleotide encoding α 1-3 galactosyltransferase.
27. The transgenic cell of claim 26, wherein said polynucleotide for expression comprises a stop codon.
28. The transgenic cell of any of claims 21, or 23-27, which is an embryonic stem cell, an ovum, a primordial germ cell, a spermatozoon, or a zygote.
29. The transgenic cell of any of claims 21, or 23-27, which expresses said polynucleotide for expression.
30. The cell of claim 29, wherein said polynucleotide for expression encodes a Type I fucosyltransferase, a Type II fucosyltransferase, an α 2-3 sialyltransferase, or an α 2-6 sialyltransferase, and wherein said cell produces said protein.
31. The transgenic cell of any of claims 21 or 23-30, wherein said cell produces a heterogenic complement regulatory protein (CRP).

32. The transgenic cell of claim 31, wherein said CRP is human and wherein said cell is nonhuman.
33. An embryo consisting essentially of transgenic cells according to any of claims 21 or 23-32.
- 5 34. An organ consisting essentially of transgenic cells according to any of claims 21 or 23-32.
35. The organ of claim 34, which is a lung, a heart, a liver, a pancreas, a stomach, an intestine, a kidney, or skin.
- 10 36. A transgenic animal consisting essentially of transgenic cells according to any of claims 21 or 23-32.
37. The transgenic animal of claim 30, which is a cattle, a mouse, a pig, a cat or a dog.
- 15 38. A transgenic knockout animal comprising a homozygous disruption in an endogenous α 1-3 galactosyltransferase gene, wherein said disruption prevents the expression of a functional α 1-3 galactosyltransferase protein.
39. The transgenic knockout animal of claim 38, wherein cells isolated from said knockout animal exhibit an increased time of survival in the presence of human serum relative to comparable cells isolated from an animal having a wild type α 1-3 galactosyltransferase gene.
- 20 40. The transgenic knockout animal of claim 38 or 39, wherein the insertion replaces DNA at the start of the coding region of said α 1-3 galactosyltransferase protein.
41. The transgenic knockout animal of claim 38 or 39, wherein the insertion replaces the promoter of said wild type α 1-3 galactosyltransferase gene.
- 25 42. The transgenic knockout animal of any of claims 38-41, which produces at least one human protein selected from the group of proteins consisting of α 1-3 galactosyltransferase, α (1,2) fucosyltransferase, and complement regulatory proteins.
43. The transgenic knockout animal of any of claim 38-42, which is a pig.

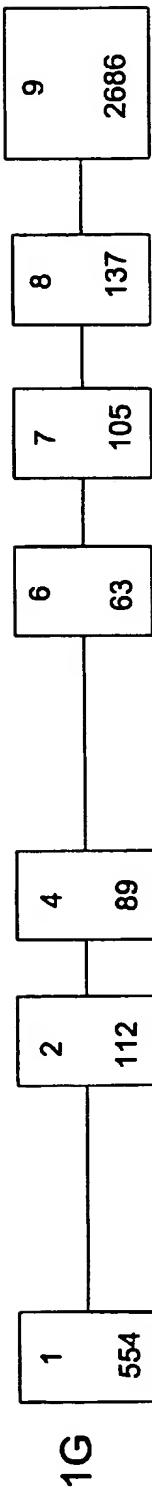
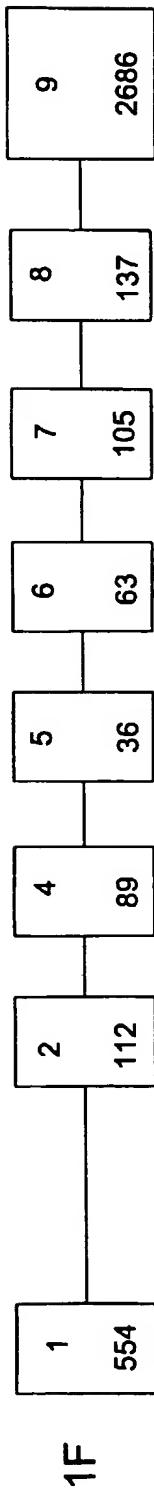
1 / 4

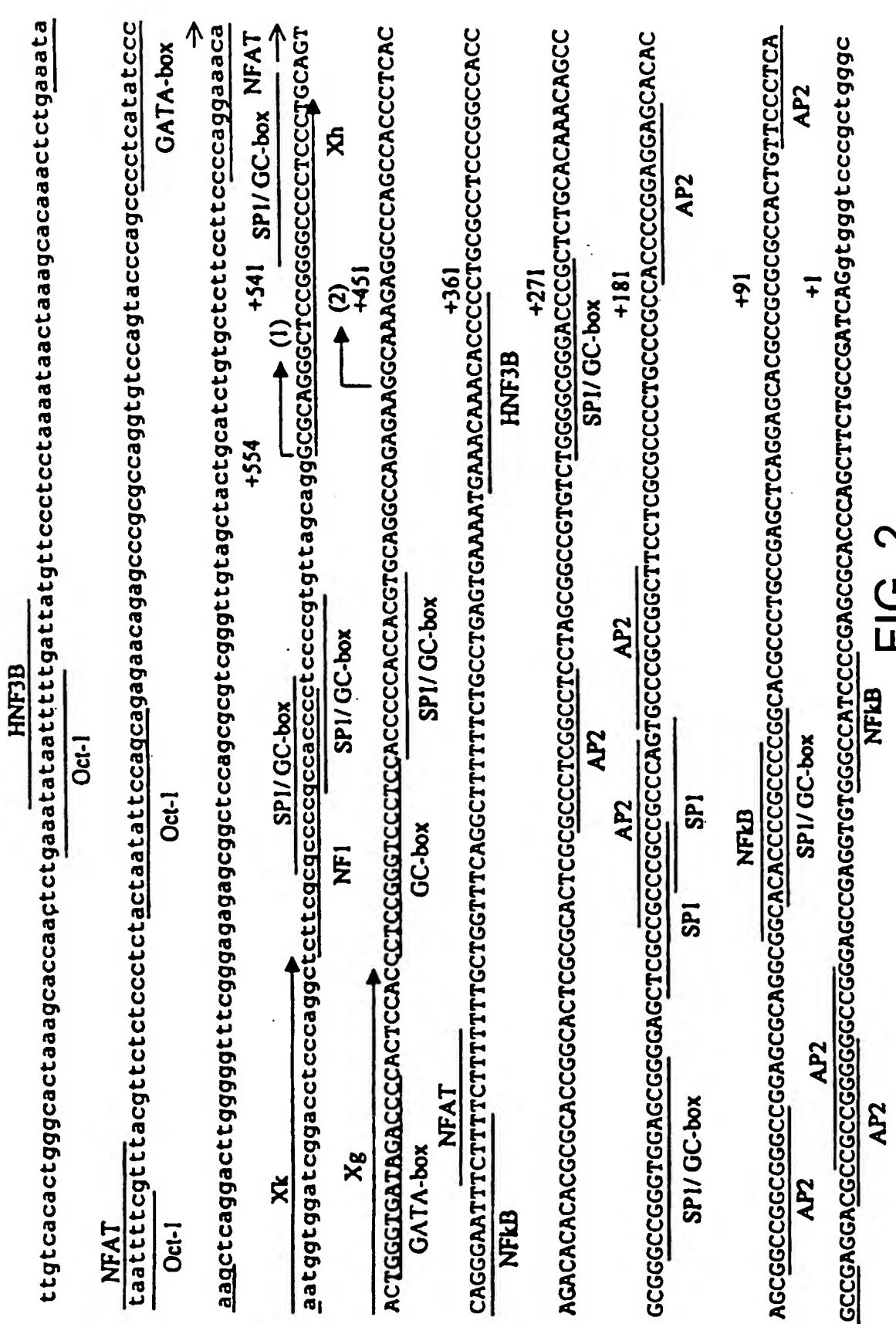


462bp 15kb 6.5kb 2.5kb 0.7kb 7kb 5.5kb

**SUBSTITUTE SHEET (RULE 26)**

2 / 4





SUBSTITUTE SHEET (RULE 26)

4 / 4

Schematic genomic organization of human $\alpha 1$, 3GT gene

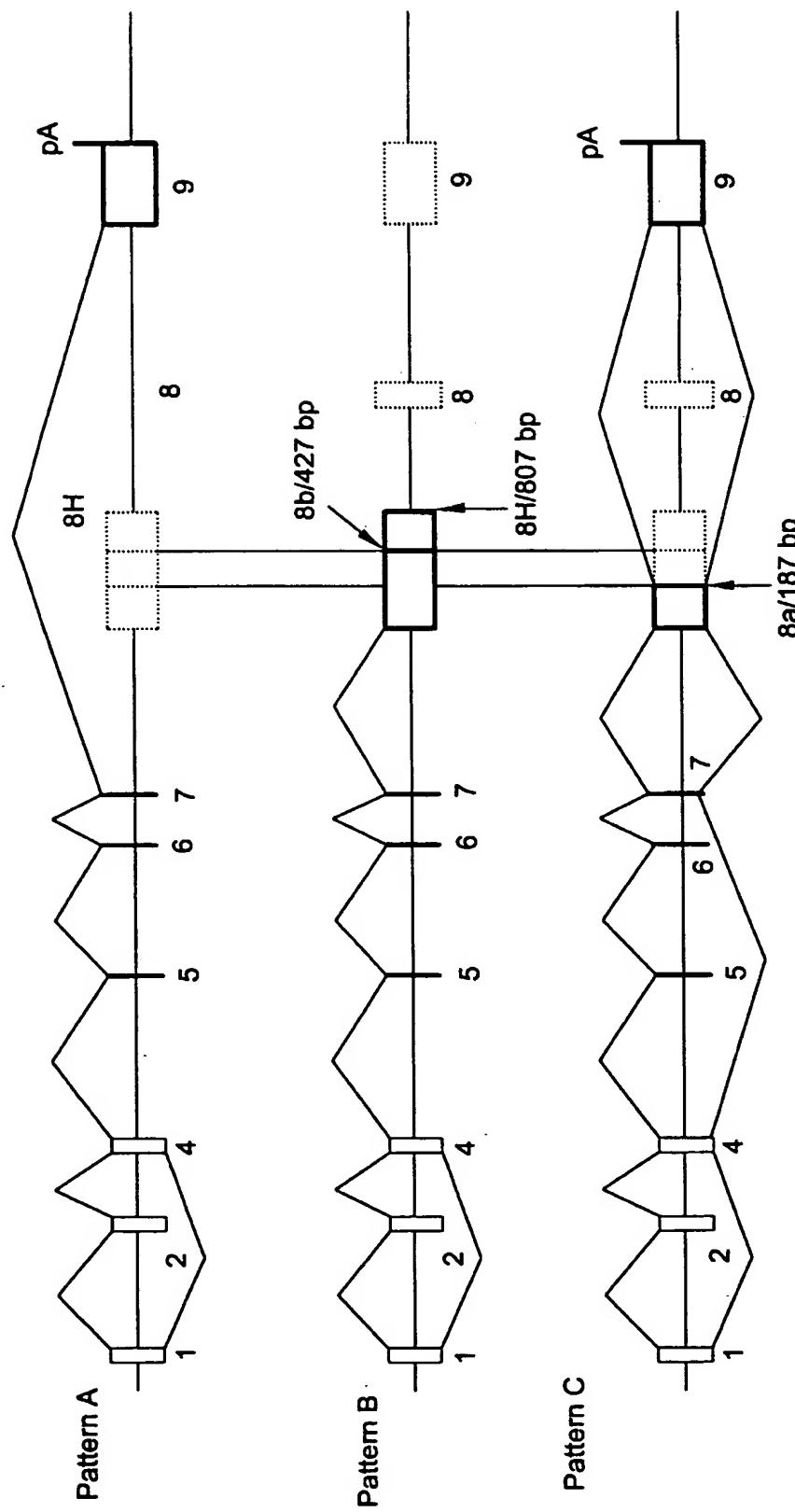


FIG. 3

SUBSTITUTE SHEET (RULE 26)

1

SEQUENCE LISTING

<110> University of Pittsburgh of the Commonwealth System of Higher Education

Koike, Chihiro

<120> alpha 1,3-galactosyltransferase gene and promoter

<130> 206780

<150> US 60/227,951

<151> 2000-09-25

<150> US 60/161,092

<151> 1999-10-22

<160> 96

<170> PatentIn version 3.0

<210> 1

<211> 1117

<212> DNA

<213> Sus scrofa

<400> 1

agatctctgt tctttcaaa tcaggatgaa acagttaaaa ttatacatca cactcaggtt 60

ctgtgccatt ttcatgtcac aattccaatg ccttaaaaata tttaagaaac taatttctta 120

gtctctgaag tcccggttg aatgatcctg gcaaaagcaa gttctgaatt ttgcagcagt 180

aaaatatagatg gtccgggacc ccaaggagtc ttgtaaaggc tgagtgaggg cagccggatg 240

tgcctacacc agctcatcag aagtgaactg ttgtcacact gggcactaaa gcaccaactc 300

tgaaatataa ttttgatta tgttccctcc taaaataact aaagcacaaa ctctgaaata 360

2

taattttcgt ttacgttctc tccctctact aatattccag cagagaacag agcccgcgcc 420
aggtgtccag tacccagccc ctcatatccg aagctcagga cttgggggtt tcgggagaga 480
gcggctccag cgcgctcggt tgttagctact gcatctgtgc tcttccttcc ccagggaaaca 540
aatggtggat cggacacctcc 2 aggctcttcg cggccggcca cccctccccg tgtagcagg 600
gcgcaggcgt ccggggccccc tccctgcagt actgggtgat agacccact ccaccctccg 660
ggtccctcca cccccaccac gtgcaggcca gagaaggcaa agaggccag ccaccctcac 720
cagggaattt ctttctttt tttgctggtt tcaggcttt ttctgcctga gtaaaaatga 780
aacaaacacc ccctgcgcct cccggccacc agacacacac gcgcaccggc actcgcgac 840
tcgcgcctc ggccctctag cggccgtgtc tggggcgaaa cccgctctgc acaaacagcc 900
gcgggcccggg tggagcgaaa agctcgccgc ccggccggca gtgcccggc gcttcctcgc 960
gccccctgccc gccaccccg aggagcacac agcggccggc gggccggagc gcaggcggca 1020
caccccgccc cggcacgccc tgccgagctc aggagcacgc cgcgcgccac tggccctca 1080
gccgaggacg ccgcgggggg gcccggagcc gaggtgt 1117

<210> 2

<211> 900

<212> DNA

<213> Sus scrofa

<400> 2

ttgtcacact gggcactaaa gcaccaactc tgaaatataa ttttgatta tgccctcc 60
taaaataact aaagcacaaa ctctgaaata taattttcgt ttacgttctc tccctctact 120
aatattccag cagagaacag agcccgcgcc aggtgtccag tacccagccc ctcatatccg 180

aagctcagga cttgggggtt tcgggagaga gcggctccag cgcgtcggt tgttagctact 240
gcatctgtgc ttttccttcc ccaggaaaca aatggtggat cgacccccc aggctttcg 300
cgccccgcca cccctccccg tgtagcagg ggcagggtt ccggggcccc tccctgcagt 360
actgggtgat agacccact ccaccctccg ggtccctcca ccccccaccac gtgcaggcca 420
gagaaggcaa agaggccag ccaccctcac cagggattt cttttttttt ttgtgtgtt 480
tcaggcttt ttctgcctga gtaaaaatga aacaaacacc ccctgcgcct cccggccacc 540
agacacacac ggcacccggc actcgccac tcgcgcctc ggcctcctag cggccgtgtc 600
tggggcggga cccgctctgc acaaacagcc gcggggccggg tggagcgggg agctcgccgc 660
ccgccccca gtgcccggcg gcttcctcgc gcccctgccc gccacccgg aggagcacac 720
agcggccggc gggccggagc gcaggccggca cacccggccc cggcacggcc tgccgagctc 780
aggagcacgc cgccgcac tttccctca gccgaggacg ccggccgggg gcccggagcc 840
gaggtgtggg ccatccccga gcgcacccag cttctgcgca tcaggtgggt cccgctgggc 900

<210> 3
<211> 1938
<212> DNA
<213> Sus scrofa

<400> 3
gaggaagggc aacatcagac ccaatggttc ctagtcagat ttgttaacca ctgagcctcg 60
atggaaactc ctgggtgctt gcttcttcaa aggaccagtt tatcttagcc cagttcctga 120
gcctccaaat gctgtgaact ttccctccca gttgaccaca gtccagctgc ctgcattcatt 180

taatgtaaa gatttccct gagtcgtac ttagtgctc tgtggtgctt ggtattgggg 240
cggtgaaccc aagagaagga aaaaacgggg tctatccacg accctgtggc cctgagaccc 300
tgttagactca gggaaagtca gaattccaa gagaaggcag cttccagcag gaagatttct 360
gtgcacatctt gtttttaaca cacacactga aaggaaatgt ttgtgaggca ttttcccaag 420
gtggacacac ctgcataacc actacctggc tcgagaaaca acatgacaag cccccccccc 480
tccccccagca gctctctgag cctcccttc ccagtctcta ccactcccac tctgacttct 540
ggcaccacag attgggtttg tctttttt tttttgtct ttttagggct acacttgggg 600
catatggaag ttcccaggct aggggtccaa ttggagctgt ggctgtggc ctacaccaca 660
gccacagcaa catgggatcc gagccgcatt tgcaacctac accacagctg gtggcaatac 720
tggatcctta acccaactgag tgaggccagg gatcgaacctt gcattctcgta acatactgg 780
cagatttgtt tctgctgagc caccatggga actccctggt tttgtctatt tttttttt 840
tttttgtctt ttttgcatt tcttggcccg ctcttgcggc atatggaggt tcccaaggcta 900
agggtccaat cgagccgtaa gccccagcct acgcccagagc cacagcaacg tgggatccga 960
gccgagtcgtg caacctacac cacagctcgcc ggcacgcctt gatcccttaa cccactgagc 1020
aaggccaggg accgaaccccg caacctcatg gttcttagtc ggattcgtta accactgcgc 1080
cacgacggga actcccggtt ttgtctattt ttgaacgtta aataaatgca agcatccagg 1140
gctgcttga ctcagtagcca tgtgtgagat ttaccctgtt gatgtcagca gctgtggctg 1200
gttccttctc acggatgtgt gtgaccctca cctggaccac acctgatctg gctgatgtatg 1260
ggccttgggg ttttccagc ttttggtccc agtcacgtc tctgtttgaa cttaaatgca 1320

5

cttgcttca ggtattaatc tggggcgaa tgactggaac atgaggtgtg gtgggttcag 1380
cttagtaca tgccagcagg gaggattca gtagttatt aagcagatct tgaagactgt 1440
ggtcaactag ctcatgcccc acaggagggg gcggtaatt tctccccc aacaggagt 1500
acaagctaaa ttaggcattcc atccgctgga agctgagggg gcagttctg gctccttct 1560
gtcagggttc ggcccttct ccttagtctg gggttctag gctctactcc caggaagt 1620
ctggggccac ttgggaacaa tgggtggggg ggctctgagc ccctacttac ttcatttccc 1680
tccttcagcc aaagccccct gtgtcctctg ttttacatag tggggtctg agaatgactt 1740
catttttttt tttttttttt ttaaagctt agctgttgcg acatttacaa atccactgct 1800
gtgaggtctc ttccaggttag gaaattgtat tttgggagca ggaggtgggt gtggggaggg 1860
ttaagcatta ttcaagccaaa gagttgggtt gggcctcagt gacctttga agttcttata 1920
gcttggcttgc 1938
ccatgcag

<210> 4

<211> 820

<212> DNA

<213> Mus musculus

<400> 4

actaaccagt gagtgtagaa agcaggaggt gtctttcct actgttagtta ggacagggcg 60
ggttggctct tcttatggac aagatggaaa aggggtgcag gtagggcaa agtgagagac 120
actcgaattt gagagacaga cagactccta acagtgaagg aaggaccaag caaaaatcaa 180
gcctgggcaa agtctcaggc actaacttttgc ctgtgttggg ttagggagg taatctcg 240
acaacttttca aaccacccctc gttccactg caaggagaca ccatcaagt 300
tttgaagatg

gcaggggaac ctctcaacaa aacacacaca caaacgtttt attatttat atttattttg 360
catgcaaagt actgtgttc attatggcat tttcatacat atgcgattgc acaaactctt 420
gaaaatcatc caagaaaacag caaagcgga aataatgtt gggggggggg gcgcggagga 480
gagagaacag agactggaga gagtgctgta ctccttgctg cggggccag gaagaggcta 540
ggagggcggg gatgtcaacg ccactagctc ctccctcagg aaggaccca gggactctta 600
ttttttagt tttgcttgc tggccacta tcggccccag aacagatctg actgcctctt 660
tcattcgccc ggaggttagat aggtgtgtct taggaggctg gagattctgg gtggagccct 720
agccctgcct ttcttagct ggctgacacc ttcccttgta gactcttctt ggaatgagaa 780
gtaccgattc tgctgaagac ctcgcgtct caggctctgg 820

<210> 5
<211> 930
<212> DNA
<213> Mus musculus

<400> 5
tgacactgaa gccacgcggg ggcttcagtg gggaggaggt gtggcgagc gcgagcgccg 60
ctattccggc ccagccctac ctcggccctt gctttgtcc tggtcactcg atcatttcct 120
ctgtatccac ttctgaactc taggctctgt cccaccctga acagtgtcgc tgcatactgtt 180
tgcttactgg ggtctccgc cacctccct cgctatccga atagctgata ttcaggcag 240
cacagggcag ggcagggcag ggcagggcga gttagggcaga tcagatcctg ggaccaccgg 300
tactaaccag tgagtgtaga aagcaggagg tgtctttcc tactgttagtt aggacaggc 360

gggttggctc ttcttatgga caagatggaa aagggtgca gtaggggca aagtgagaga 420
cactcgaatt tgagagacag acagactcct aacagtgaag gaaggaccaa gccaaaatca 480
agcctggca aagtctcagg cactaacttt gctgtgttg gtatggag gtaatctcg 540
cacaacttt caaaccacct cgccccact gcaaggagac accatcaagt gttgaagat 600
ggcagggaa cctctcaaca aaacacacac acaaacgtt tattatTTA tatttatttt 660
gcattgcaaag tactgtgtt cattatggca ttttcataca tatgcattt cacaactct 720
tgaaaatcat ccaagaaaca gcaaagcggg aaataatgtt gtgggggggg ggcgcggagg 780
agagagaaca gagactggag agagtgtgt cctccttgct gcggggccca ggaaggaggct 840
aggagggcgg ggatgtcaac gccactagct cctccctcag gaaggacccc agggactctt 900
atttttgttag ttttgcttgt ctggccact 930

<210> 6
<211> 501
<212> DNA
<213> bovine

<400> 6
cctccctgtc catcaccaac tcccgagct cactcagact catgtccatc gagtcggta 60
tgccatccag ccatctcatc ctctgtcgtc gccttctcct cttgtcccc aTCGGCACA 120
gcattcagagt ctttccaat gagtcaactc ttgcgtatgg gtggccaaag tactggagtt 180
tcagctttag catcatcccc tccaaagaaa tcccagcggc cgagtccggg gcgggacccg 240
ctctgcacaa acaccggggg ccgggcccag ctgggagcgt cgagcccgct gcccagcgcc 300
cgccggctcc ctgcgcggcc tggccggccgc cccggaggag cgcccgccgg ccggccgacg 360

ggagcgccgc ggcacacccc gccccggcac gcccgccggg ctcgggagga ggca ggcgc 420
cgactgttcc ggcagccgag gacgccgccc gggagccgag gcgcggcca gcccccagcg 480
cgcccagctt ctgcggatca g 501

<210> 7
<211> 3976
<212> DNA
<213> Sus scrofa

<220>
<221> misc_feature
<222> (580)..(580)
<223> "n" is a gap of from about 600 to about 800 nucleotides

<220>
<221> promoter
<222> (1863)..(2992)
<223> fragments and derivatives of this region have promoter activity.

<220>
<221> 5'UTR
<222> (2463)..(3016)
<223> Untranslated exon 1 runs from about nucleotide 2436 to about nucleotide

<220>
<221> Intron
<222> (1)..(2462)

<220>
<221> Intron

<222> (3017)..(3976)

<400> 7

aggcctaaac ctagaactcc tgaccctgaa gctaaggaat ataatcttga aggtgttttc 60
cagtca tag aataaacacag agtttccaca catgcgtggg tctcttctta gggtgcttat 120
tctgttccat tggccaata aaccatcctg gcgcta atgc tatactgagt tcactgcgtt 180
tcatggtctg tcttggtatac tggtggaaaca agagcccaac tctccctcc ctgcttgtc 240
aagactgcct tggttatatac tggcccttc ccgctgctgt ccaaattta agaata gctg 300
gccaa gctcc cccaaaactc tggcatttgttggatgt ttataggtt atgc atggag 360
aattgttgcc ttcgtgatgc tggatgttgc cagtgctcac tcgggggtct cttccttcc 420
acctaaagac ttctgc acat ggttctgctt gggtcactct tccccaa gccc ttccac tagt 480
gaactcctcc tcctcctggt ctcagggtct cctgc accct tatttcttcc ttagagccct 540
gatcacaatg gtcctgaaat cactcattgc gtgggtcttn gtgacagata gtaggtccca 600
gtaaaatatct gtaaaagaa tgaaggaat ttaggttagga aggtttcg gacctggagc 660
accttggcca tagtttagagg gatggtgacc agaggtactt aacttgcctg tgccttggct 720
ttcttc tacaaaacccggga tgtgatcaga atgtgtataa gatgaatgt gctc agttag 780
gccgtgaggc aagtggagca aagcctggca agggatcaga gctacttggt tacctgcct 840
gcccttctgc tcagtgaatc ttca gtcctg cactcctgtg atgctcctgg aggctccaaac 900
actcttcccc cagcagtgtatcc cccgttca cttccacccctt cctatgaact agtc accttta 960
tttctactca gcatatgaca caaatgagtc tcaggaagaa tgactcataa ggcctt aaac 1020
ctagaactcc tgaccctgaa gctaaggaat ataatcttga aggtgttttc cagtc tag 1080

aattgctagt tagatttggg gagctacata gttctcaaaaa gaaaacaaaaa cttccggacc 1140
cgccgtgtta atttgaatta ttttatctt attgttactg aaataggtat aaacctagaa 1200
ctaagaatga agtcctcatg ctccctagctc tgcacacacca ccatgataacc aaagcaaatc 1260
ttttaagtag gtgcaattac agccacaaaa ccaataaaaat ccaaattagc aacgttaaat 1320
ttatgcaact gatgacatgg tgctgaaatc aaacctcttg cattgagtct aatggtagca 1380
gagtgatgtt ttacatgtt tcattccctg tgtcatcatc ttttgatttt gatcctgatg 1440
agctatcact tcagccatgg tcagaattac cgctcataatt ttcactaaaa aaaaaaccca 1500
aaaaacacat ttattatcca atttgcgtttt ctgagcaatt taaacactgg atcctcaagt 1560
gcaataatga caactggaa atacttgct aacatcactc cttgtgtatt tatttactgc 1620
atcattaaag accttagtgca agtgaggtaa ccgatgacaa taatggcgca gtttatgctt 1680
ttgcaaagga tccattgttc ggattgtcat ggagctcctc attcctgagc taccctgtgg 1740
ggctgatgat tcaactctcc caccctttag tccactgaac ccatcaggaa agttcattat 1800
ccccagctcc aagatgtcac ttggctccct gcagcctctc tgcaaccgtc aagtattcaa 1860
tcagatctct gttttttca aatcaggatg aaacagttaa aattatacat cacactcagg 1920
ttctgtgcca ttttcatgtc acaattccaa tgcctaaaa tatttaagaa actaatttct 1980
tagtctctga agtcccgtgg tgaatgatcc tggcaaaagc aagttctgaa ttttgcagca 2040
gtaaaataga tggtccggga ccccaaggag tcttgtaaag gctgagttag ggcagccggaa 2100
tgtgcctaca ccagctcatc agaagtgaac tgggtgtcaca ctggggacta aagcaccaac 2160
tctgaaatat aatttttgat tatgttccct cctaaaataaa ctaaagcaca aactctgaaa 2220

tataatttc gtttacgttc tctccctcta ctaatattcc agcagagaac agagcccg 2280
ccaggtgtcc agtacccagc ccctcatatc cgaagctca gacttggggg ttccgggaga 2340
gagcggctcc agcgcgtcgg gttgtagcta ctgcatactgt gctttccctt ccccaggaaa 2400
caaatggtgg atcggacctc ccaggcttt cgccccccgc caccctccc cgtgttagca 2460
gggcgcaggc ctccggggcc cctccctgca gtactgggtg atagacccca ctccaccctc 2520
cgggtccctc cacccccacc acgtgcaggc cagagaaggc aaagaggccc agccaccctc 2580
accaggaat ttctttctt ttttgctgg tttcaggctt ttttctgcct gagtgaaaat 2640
gaaacaaaca cccccctgcgc ctcccgccca ccagacacac acgcgcacccg gcactcg 2700
actcgcccc tcggcctcct agcggccgtg tctggggcgg gacccgctct gcacaaacag 2760
ccgcgggccc ggtggagcgg ggagctcgcc gcccggccca cagtggccgc cggcttcctc 2820
gcgcacctgc ccgcacccca ggaggagcac acagcggccc gcggggccgaa gcgcaggcgg 2880
cacacccgc cccggcacgc cctgccgagc tcaggagcac gccgcgcgcc actgttcct 2940
cagccgagga cggccggggg gggccgggag ccgaggtgtg ggccatcccc gagcgcaccc 3000
agcttctgcc gatcagggtgg gtcccgctgg gcgcgtcccc agcccttggaa ggccgcgag 3060
ccgcggccgc cggggctgc gggcgccgtg gaggcagcgc ggggagagga caggccaccc 3120
cgccggccct gcctgttgc tgccctgccg tgtccccgt tttgttctcg tcgttacctc 3180
tgtgctcaac tctgaccccg tctctgtccc catttgtcg ggcctgaggg gctgcgggct 3240
tccacgggtt cgcggatg gaggcgggag aggggaggct cggggcgcgc agaggaggag 3300
gactgccccg gaagtctcgaa aaggaggag gggctgtct cccaatgtgg ggcaggagg 3360

gcggaggcct ccctcgcccg ggacttaggtg ggaagaggat gcctccgcaa gagggAACCT 3420
gagagtgaag tggggggcac agaaacccctg aacgcacaga gagggagaag tcggggAACT 3480
cagagagcgg aggaccgaac ccgaaACCCG gccggggaa acttttggAAC cccgaaACTT 3540
tggcggcgaa aaaggccgct gtatcggtg acaggaagca aagggtcctt cagactttAA 3600
gccacacgtt ccaggaggGA gggaggcgcg gagaccgtct gcgggcgcgg ctccccc 3660
cagaaaAGAC aagagacCCG gacgggtgct ttgtggTT tgcttgcgt cgTTGCCCT 3720
cctcttggcc cctgagcggg ccttgtcgcc ttgttcttGT gcttgaaAT gggtgggtct 3780
cgagcgctg gacgtgcggg gacGGGGGG gtggggcga ggaggagtCG gggccgggac 3840
gcctccTAGC tgccaaACCC tttccaggG agaatccgtt tccacaaACC tggaaatAGAG 3900
agactgctgg aagtaaggaa atgccaagtG cgaagaggTT gtgtgtgtGT gtggtgggg 3960
gggatgtgGA tgctt 3976

<210> 8
<211> 8989
<212> DNA
<213> Sus scrofa

<220>
<221> Intron
<222> (1)..(4731)

<220>
<221> misc_feature
<222> (4732)..(4814)
<223> untranslated exon 1A found in some transcripts

<220>
<221> Intron
<222> (4815)..(5241)

<220>
<221> misc_feature
<222> (5242)..(5353)
<223> untranslated exon 2 found in some transcripts

<220>
<221> Intron
<222> (5354)..(8989)

<400> 8
aaaatctgat tttgatctga ttggctagt ttatcacagt ccatccttac ctggtcaaat 60
tcacatactt ctgctgcctg cctggctcct gtaggcttc actcagcatt aattcagcaa 120
atatttactg aacatctgat agatgtcaaa tactgttcca ggtaccagga aagcccagaa 180
gtgaccaaga cagaagacaa gtgctccctc ccacccccc aagagcttgg gttctagtgg 240
aatctggttc atgaccctct tcttggtctg cctccgttag catccccagc ttggtctgac 300
ttcaccacca ccaggggtgt acaaggctga ggtggacag actcacagaa agaccta 360
tttgtcttcc attccagggc tgctgactca taccatacga ctctgttaagt ttcttccctg 420
atcttcagtt ccctttctta taacttgggg ctgtaatat ttcacctact tagcctctat 480
tttatgtggc ttttgtggat ggcagtgggc tctaaacggg gcgtgggtgt gaccttgacg 540
gaagatgagc ttatcacgtg ttcaaaaagc agtcctgctt tgaggcaggg agctgactta 600
cctgactttg agttctctc tgctgaggaa agagtgagaa cttctgtggg gggtcgaaaa 660

caagggtacc ccctggcacc tactgccaa ttgtgaataa ggagcaggtg cctctttctc 720
acacctccatct ggggtacttg gcctgaggaa ggggtgagaa ggaccaagag aggttaggaa 780
tagagcggtt tccttgggtg gggaaatcct ccagtcacct gtgctggtgc tcaagcccag 840
gctgtcatca gtacccgggc ctcgccttc cgtggagcg cctcacatct ccccagctgt 900
caacaaagcc agcttcttc ttctcttagga agagtctgac ctatagagct tgaaggactg 960
acatgagccc cagagaggga cttcctggtg tgcaggagga gggctgaggc tcaggatgga 1020
tgcttgcaga ggcaggagtg cttcagcatg gctttggtgg agtctgtcct ggagttacct 1080
ggggcagagg cagatctcaa gatgattagc aatgtactgg cctggaaaga gtcatcatga 1140
tttcattttt ccagctcttc tcaaggaat agacttatac atgcaacctc tcttgactgc 1200
cgttatttat tatgtggct tttgccaaga tcgttcagc tctgatactc acaggcgtgt 1260
gtggggggca gtacttaaca gtaacggaaa cgtcgtgcc ggaacccttc cctccgtacc 1320
tttccccacc tgcagggta catggtaaaa atgactattt gatacacaaaa tgtaaactcc 1380
aaggagctgc agcctcgat taatagaaca gcagagacgg acaatgattt agcacctcaa 1440
gcactttcc gggcgtgtct ctttacttct tgcaatattt ggtataacgt atctctagac 1500
acttaccatg tgccagctac catccagctg ctgttggtcc cattgtgcag ccgtagaaac 1560
agagacacag agaggttaag cacattgccc aggatcgcat atgggcaggc ctgggactcg 1620
aactccggca gcctgggccc agagtccaca ttcataacca cggtgctcta ggcccctcac 1680
ccaccccgag cgggtgggat tataattatc ctcaccacac ggaagaggaa accaactaaa 1740
ctgctccatc actcacaagt gacagcaaga atgtcttata cctgccttaa acgtatTTAG 1800

15

gattaaaagt gacagctgca acctttgtat ctgttagcact ttttgc当地 aacacttaat 1860
cctccctctc ccacagggtg ggaatccgga cctttgtgtt tctcagctgg aaggggtctg 1920
gggcatgaag ccgggaccct tcacacctgg gctgcagctg ctgagccgca gctccaaggc 1980
cctgcactcc tctgcagggg acatggcaga tggacaggct ctgaatgctg gctgtcatct 2040
gacaggccta tggactgtta gggcttggaa gggccttggg gaacatttagtg tgatgagatt 2100
agtcggcctg gctgggctgg gaaacgtgcc aaactcctac ctggatggcc actggcctcc 2160
tttgatcagc agacctgagg ctcacttgct acagttccct gcctctccat gaaggaatgg 2220
ccggaagtac atgtttcctt gtttgagag tctggcatac aggatgttc ggagaaggag 2280
gaaggtcatg tcggatcctc tggaagttga attttctgccc ttccaagttt gcatactctg 2340
tcgtgctctg attcatgaac ctggagcctc taattccacg aacctgttagg gtgttcccc 2400
gaggcagctc aggaggaagg gcagcatcag acccaccagc cgccaacttt gagcaagtca 2460
cagaggctcc cagtgcctcc ctcccttccc tgacccgggg cgggtgagcc tgaggattt 2520
ctgagttaaa ggagagagggc tgctttgtaa actggaaagggt ggcaaccatg atgggtgctt 2580
gtttttttt gtgttgttg ttttgtttt ttgtctttt gcctttcta gggccgctcc 2640
tgcagcatac ggaggttccc agcaggctag gggtaagtt ggagctgttag ctgccagcct 2700
acgccagagc cacagcaacg tggatctga gccgcgtctg caacctacac cgcaatcac 2760
ggcaacactg gatccttaac ccactgagcg aggccaggga ttggacccgc aacctcatgg 2820
ttccttagtca gatttgtttaa ccactgagcc tcgatggaa ctccctgggtg cttgcttctt 2880
gaaaggacca gtttatctta gcccagttcc tgagcctcca aatgctgtga actttccctc 2940

ccagttgacc acagtccagc tgcctgcac attaatgtg aaagatttc cctgagtccg 3000
tacttagtg ctctgtggtg cttggatttg gggcgttgaa cccaagagaa ggaaaaaaacg 3060
gggtctatcc acgaccctgt ggccctgaga ccctgtagac tcagggaaag tcagaattcc 3120
caagagaagg cagttccag caggaagatt tctgtgcac tttgtttta acacacacac 3180
tgaaaggaa tggggatgag gcattttccc aaggtggaca cacctgcata accactacct 3240
ggctcgagaa acaacatgac aagccccccc ccctccccca gcagctctct gagcctcccc 3300
ttcccagtct ctaccactcc cactctgact tctggcacca cagattggtt ttgtctttt 3360
tttttttttg tctttttagg gctacacttg gggcatatgg aagttcccaag gctaggggtc 3420
caattggagc tgtggctgtt ggcctacacc acagccacag caacatggga tccgagccgc 3480
atctgcaacc tacaccacag ctggggcaa tactggatcc ttaacccact gagtgaggcc 3540
agggatcgaa ctgcattct cgtacatact ggtcagattt gtttctgctg agccaccatg 3600
ggaactccct ggtttgtct atttttttt tttttttgt ctttttgcc atttcttggg 3660
ccgctttgc ggcatatgga ggttcccaagg ctaagggtcc aatcgagcc gtagccccag 3720
cctacggccag agccacagca acgtgggatc cgagccgagt ctgcaaccta caccacagct 3780
cgccggcaacg ccagatccct taacccactg agcaaggcca gggaccgaac ccgcaacctc 3840
atggttctta gtcggattcg ttaacccactg cgccacgacg ggaactcccg gtttgtcta 3900
ttttgaacg taaaataaaat gcaagcatcc agggctgctt tgactcagta ccatgtgtga 3960
gatttaccct gtgatgtca gcagctgtgg ctggttcctt ctcacggatg tgggtgaccc 4020
tcacctggac cacacctgat ctggctgatg atgggccttg gggttttcc agctttggt 4080

cccaggcac gtctctgttt gaacttaaat gcacttgctt tcaggtatta atctggggcg 4140
gaatgactgg aacatgaggt gtgggggtt cagcttagt acatgccagc agggaggatt 4200
tcagtagttt attaaggcaga tcttgaagac tgtggtcaac tagctcatgc cccacaggag 4260
ggggcggtga atttcttccc cagaacagga gtgacaagct aaattaggca tccatccgct 4320
ggaagctgag gggcagttc ttggctcctt tctgtcaggt ttcggccct ttccttagt 4380
ctggggtttc taggctctac tcccaggaag tgtctgggc cacttggaa caatgggtgg 4440
gggggctctg agcccctact tacttcattt ccctccttca gccaaagccc cctgtgtcct 4500
ctgttttaca tagtggggtt ctgagaatga cttcattttt tttttttttt ttttaaagc 4560
tttagctgtt gcgacattta caaatccact gctgtgaggt ctcttccagg tagaaattt 4620
tattttggga gcaggaggtg ggtgtggga gggttaagca ttattcagcc aaagagttgg 4680
gttgggcctc agtgaccttt tgaagttctt atagcttggc ttgcctatgc ggagatctca 4740
gaacattcta taaaaatagt gttcaaacag aacaacttct gaagcctaaa ggatgcgaac 4800
aaggaggctcg gaaggtagca tttcaacggg agttttgagg atgcttcctt ttagccaccc 4860
ctctccattt tctgccccct tctttttaaa ttctccattt gctgtccctg ctatgttca 4920
tttgggtgg tttgggtca gaatggttct cattttcgcc gaggagtggg tgatgtggc 4980
ggcctgtgtg tctctccaa gggtggtggc tgtccctcct ccaccaccag gcctagttt 5040
gacctgttagt ttgccttagt gaaggaggcc gggccgatcc tggggccggag agagacgtct 5100
ctgccttggc atgcagctct gagtcaaacag gcctgataaa cagcccaattt cccaggcgaa 5160
gcaaggagga acaaggcccc tggctgctgt gggatccgctc tgcgccttc ttcgtgaaac 5220

18

cgctgttat tctttgaca ggagttggaa cgcaagcacct tcccttcctc ccagccctgc 5280
ctccttctgc agagcagagc tcactagaac ttgtttcgcc ttttactctg gggggagaga 5340
agcagaggat gaggtacgtg aaacgttgaa atgatttacc tccgcttgc tggggtcacc 5400
gggggggtgg gtatcatgag ctggctgcag cgtggagaga ggagcccccc tctccccctg 5460
acttcttgct gctcccccca gttgttctga aagaagacaa agtcctccag tccccggcat 5520
cgatcttagg agtgggagct ggcaggatgc tggctcagtc actgttggtt ctgcttcgt 5580
tggctgcccgc gcaggacctc acggggtgtg gctacagcct ggggttctct gtgtgggcca 5640
cacagtgcac ttgtgggccc aggaggacga gtctcaggcc cgggacctgt gctggggcg 5700
gacatagtgc cctctcaggc cagcaccgat cttcatgta cctcgcctta tttctttgg 5760
aaaaactctt gcaccatgat ttctgagcca ggcagcaagg agaagctggc tggatccagg 5820
cttcagattt ttgaaggggaa ttcaagaaag gggcctacaa gatgtccctc cgagaacagg 5880
tctgtatgg ctggagcgac agctgtgaaa aaaataagtg gaaagagcct tcggtgcgg 5940
actccccccc cacccctgcc ccccaaatta taccatgttt cttccaacag ggagcatttc 6000
cctgtaatgc aagccaattt aaattcttga gggtgcacat tttggttta tttcaactga 6060
ttattagtgt agaggagtat aagataacat ttctttaaaa accatcaaca caaaccatc 6120
actcgtgatt caattgttta ggagaggagg gaactccgccc tcgtataccca aatacagtct 6180
gctctcggtg cagcgtgcag tcccagcaag gccctctcct cgaactcaca cagctttgt 6240
ctccagcggc ttcttccca tgtctggct aggctggct ttcttagtaa ccccaaaggc 6300
ggagaatcaa attcacagat ttttttttc tggatattta gatcttgtat tttaagccac 6360

actatttata aggctcagag atacatcaa actctgacta gggcttctta taaaagtgtat 6420
atctggaaag aaggctggtc tttaacagag taagggtcag accccccc ttcccatcaa 6480
tgactccagg aatgctctgg aagactgaag tggaggcaaa gaaggacttg aatttgcattg 6540
acctgatctt gaatccaggc taaattttc ctggctgtgc gccttaggt gggtcattta 6600
cctccctaa ttctcaggtg gctcacttca tcatttattc ttttactgag gcagagaggt 6660
ccctctacca ccaggttgaa tgagctcagt gacctctgaa aactccaaag tgctgcacag 6720
atcaagggtgg tatgaggttag aagaggaagg gaaaaaggaa tgagttaggat caaagaaaga 6780
aggagtggaa agaagcagag tggagagaca gagccaacac aaggatctgg gtaccacttc 6840
tggattaggg tcagggctta gaagatgaca ttgatggttt ggtcttttc actacacaga 6900
gaatagagct gaccattaga cttggcccg agccagtcat tgtgaaagaa atcaatattc 6960
agattatcat gacaactacc atttgtgtaa ttttaattca caggatcact tttctggcc 7020
cacgaggttg aaataagaat ggctggtcag attgactggg gcggtccgac tggcctgtgc 7080
ttgagagttt accatgagct ccctgccatc tagcgtgtat gtcacccaga ctttaactc 7140
accatctgga ctgaccctcg agaacttgat gccatttgag agcacccaaag gggtccagag 7200
gaccttatca aatcctctga ctcctctgtg caggctgttgc cccagcttat actccttccc 7260
atccaacgtg atgttccttt ggcaatttgc tttgccaccc tgccaaccac tgctccaaag 7320
tagggatgct tttggaggtt cccttccaat tcagcaaagc caagcaccac atctgaggct 7380
ctgccttgcc tgtcttgac ctccaggccc gtgatggtgc agccccgagga gatgatttcc 7440
actcccagtg ttgttcagcc cgaggagatg atttccaatt cccagtttgtt ctgcttgca 7500

20

ctggaaattt tccatgttcc ttgccccaa ggggagttct ccaaacacag atcttgtaac	7560
tgaaaaccatg aggaaagctt ggggtgtgta ggtgctccag gtccttcaaa cgccccatct	7620
tttggcagtt tcttgctcag gtgggtccag ccagagtcct ggagaattca gcttttgat	7680
cctggctgga gtggggggtg caccaccagg tgattgtgag gtctggatcg tgacctgtga	7740
gcagggagcc aagtagcatc atgttcagct ctttcctt gggatcaaag tgagaggctc	7800
caaggagctc agcaaggctc acctggatgg ggcaggttgc tcctaggacc caggttagtg	7860
cggggagcag ggtcagtacc tgggctccac ctgcagcccc aggacaggca cccaggctgg	7920
aacgattccc ccaggcaggg gcagcacctc acctggagga agcattggg cttgcccac	7980
tccacacccc aggccgcct gggggcctga cccggaggct tctgggtgaa gtggcctgag	8040
ggctcaacac attttgtggg caatcctatac tctttttta ttttatttt ttatTTTT	8100
gctttttagg gccgtaccccg ctgcatatacg aagttccttg gctaggggtc aaatcgagc	8160
tacagctgcc agcctacacc acagccacag caacacagga tccaagccgc gtctgtgacc	8220
tacaccacag ctcatggcaa tgccggatcc ttaacccact gagcgaggcc agggatcgaa	8280
cccgcaacct catggttcct agtcagattc attccgttg cgtcatgacg gaaactctgg	8340
caatcctatac tttgatcac cacttctagg aatctgtggc cactgcagca agttgagctc	8400
cagtgaacct gtcctataa aaggagcctt cagctctgtg gtcgccttct catacaggctc	8460
ttggctcatt caggggaagt taagcccaca ggacatgttt caaaggacgg gaaatgcact	8520
gggttttagc acagtctgca cgaggcccg gagtgggggt gcaagtggtt tctttggaa	8580
accgctgcag gggctgagtt gtgggagtgg cccaggagca gagagaaaatg gcaaacgcct	8640

21

tggcaggagg gcctgtggga tggtgggagg gctcaggtgg aactgggccc gctgggttca 8700
cctgatcctc tgagggctgg ggcccaggtg gtgctgaggt gttacactc tcccttataa 8760
gacaggatgc tagtgctctc taggctctaa tcctgtgctc tccctttcc atgagaaaatg 8820
tagaagcaac ccccactttt cctatttgtt gggtaagata gtcaaccacc aatcttgaga 8880
attagagagt tttgaaaatt ctgtgacaaa cacatccgtg aaggccttt agaccacatg 8940
ggctgc当地 tgc当地 taatccagag agaaaaataa aattgttt 8989

<210> 9
<211> 240
<212> DNA
<213> Sus scrofa

<220>
<221> Intron
<222> (1)..(29)

<220>
<221> misc_feature
<222> (30)..(118)

<220>
<221> Intron
<222> (119)..(240)

<220>
<221> 5'UTR
<222> (30)..(38)

<220>
<221> misc_feature
<222> (39)..(41)
<223> This "atg" is the translation start codon

<400> 9
aattttccct tctcctttc ttttcccagg agaaaataat gaatgtcaaa ggaagagtgg 60

ttctgtcaat gctgcttgta ctaactgtaa tggttgtgtt ttgggaatac atcaacaggt 120

aattatgaaa catgatgaaa ttagtggat gaaagtctcc tctaattctcc tagttatcag 180

ccaagtcacc agcttgcatt aaaagtagga ttcactgaca ccgtaaagaa agcattccag 240

<210> 10
<211> 2685
<212> DNA
<213> Sus scrofa

<220>
<221> Intron
<222> (1)..(2140)

<220>
<221> misc_feature
<222> (2141)..(2176)
<223> This region defines exon 5

<220>
<221> Intron
<222> (2177)..(2685)

<400> 10
aagcttttaa ggactctaag ctttcatttt tctttttttt tttccttatct tcgacttggt 60

tgcttaggaag cttagagcaa agtattgtgc ttaaatgctt gcattttcct tggccttcatt 120

tttttttaaa acatTTTTC ttatTTAAGT atagctgatt tataCTGACc ttcatCTGAT 180

23

atgatttatac ccctggtggtt aaatccctggc ttttggtaga tgccatggga tcttggcaat 240
ttgctcaaac tcattttgcc aatatcttag ctatgaagta aaaataaaagt taaagatttt 300
gttctcacag agtggctggg atgaccaaag tcatgtgaaa acacccgagt gactaaaatg 360
tttctctgtt tcgttttgtt ttgtttgtat tcttgttattt tttccattt tatcgtaacc 420
acactttctt cataagccat ttcaaggact tcctgaaagt agatggactt taagtttctt 480
ggacttccag ttgtggcgca gtgcaaacaa atctgactag tatccatgag gatgcacatt 540
cgatccctgg ctttgcctcag tgggttaagg atctggtgct gctgtgacct gtgggtttagg 600
tcacagaggc ggctcagatt ccaagttgct gtggctgtgg cgtaggccgg cagctacagc 660
tccaatttata ccccttagcct gggaaacttcc acatgcccga gggtgcaacc cccaaagata 720
aatgaataaa taaataaaata tgcgacccctc ctttcttggg gcccatttgcattt gttttctct 780
ctgttaggca cactcttgct aatccctctt cactgggcct cctatgtatc ctccagaact 840
cagctaaaac atcatccccct cccctggggc gccttcgagg tcttcctgtt aagtgtccct 900
atgctttctt ggagtttga agtcctataa tgatgtgttt atcaaaatag ggtccaccct 960
ccctgccagc ttctttacac cacagacaca tggtgtctgt ttcagtcac actgtatgtc 1020
tggcacttga catgtAACGC atgctcagca ggtattttgtt gaatgaatgg aggccgtctg 1080
ctagagtcgt catatattta ctgatcccgat cttgttaggat ggtctcactg cttttgttag 1140
cttaagaagt acctttttt tttttttttt tttaatggcc acacccatgg catatagaaa 1200
ttccacgaag gaaggaagaa agaaagaaag aaagaaggaa attcctgggt cagggattga 1260
atccaagcca caggtgcaac ctgagctgca gttgcggcaa caccacatct tttaacccac 1320

tgtgctggc cagggatcat acctgtcat ctacagcgac ccaagccacg gcagtcagat 1380
tcttttctg ccttcttc tttctttct tttttttt tttttttt ttgtcttt 1440
tgcccttct aggtgcggca tatggagggtt cccaggctag gtgtcgaatc agagctgttag 1500
acgccggcct aaaccacggc cacagcaaca caggatccaa gccttgtctg tgacctacac 1560
cacagctcaa cgcaacgtt ggatccttaa cccgttgagc gaggccaggg attgaaccccg 1620
caacctcatg gttcttagtt ggattcgtaa accactgagc catgatgggaa actcctgcag 1680
tcagattctt aacccaccat gccacagcag gaactcctag aagtgcctt tgaggctact 1740
ctgttagacag ctttgagcca gcgaggcaag acctgtttt ctggaggaag ataaatcctg 1800
ggtagggat gggtagggctg tggcttcct gggacccatc tctggagcct ctctccctca 1860
gcaaagccac cttggacaat aagagctgcc atctatTTT ttttctta aactaagatt 1920
tgatatttc cagagacctc cctcccacccg ttgcgtctga gtaattctga aatgacgaga 1980
gccccgtgat atcattttt cgatctcgaa ggtggaaacc tgggagtagc cacaacccag 2040
gctctcagct cagcctaggg tttcaatgtat aatgattgca aaatagctt tctctgcgtt 2100
ccaagtaaca tggatgttt ttatTTTcat ttgttttag cccagaaggt tctttgttct 2160
ggatatacca gtcaaagtaa gtgcttgaa ttccaaatat ctctaggta ccttccatgt 2220
gaccctggtg gcccatacgat ccattcttaa catggcaggt ggtgacgcac ttgtggct 2280
aggtagggaa gagggatggg gttccagggg tctgagctgt acttctccag cccctagact 2340
tgccttcta gagcatgagt tgtgttttc ctttgcttct catcaagtat ctatctttt 2400
aagtgtatgtt gttggagaa cattcctgcc ttgctcataa aaaagaatca gagtagat 2460

25

tatccattat gctacctact acatgtggta taaagaccct tgcccagaaa ttttgc
tatccattat gctacctact acatgtggta taaagaccct tgcccagaaa ttttgc
ataaaaggatt aggaagaaaag gctgggtgtc ctgataaaact aagtgtgtgt attattatta 2520
tttaatatta ttactaatac tgggtgattt aaggactcc taaggccttc aattttcct 2640
ttttctttt ttttcccta atcttccgac ctttggtttg cctaa 2685

<210> 11
<211> 180
<212> DNA
<213> Sus scrofa

<220>
<221> Intron
<222> (1)..(37)

<220>
<221> misc_feature
<222> (38)..(100)
<223> This region defines exon 6

<220>
<221> Intron
<222> (101)..(180)

<400> 11
tttctaaaaa atgttgtca tcttttcat ttcttagaaa cccagaagtt ggcagcagtg 60
ctcagagggg ctggtggtt ccgagctgtt ttaacaatgg gtaagactgg gaaacggcca 120
tctgtgtatc tgctcaaggc tgttagagtcc aaataaaaatg gttcacagc catgaccc 180

<210> 12
<211> 242

<212> DNA

<213> Sus scrofa

<220>

<221> Intron

<222> (1)..(100)

<220>

<221> misc_feature

<222> (101)..(205)

<223> This region defines exon 7

<220>

<221> Intron

<222> (206)..(242)

<400> 12

atgacaccttct ccagtcgcgt cgtccttctg gcttatttggaa cattctggca catgggtcac 60

cctccctgcc ttcctcagct tgtttccgt ttgtacgttag gactcacagt taccacgaag 120

aagaagacgc tataggcaac gaaaaggaac aaagaaaaaga agacaacaga ggagagcttc 180

cgcttagtgga ctggtttaat cctgagtaag aaaagaagcg ttgccctatt tcagtaaatc 240

ca 242

<210> 13

<211> 720

<212> DNA

<213> Sus scrofa

<220>

<221> Intron

<222> (1)..(257)

<220>
<221> misc_feature
<222> (258)..(394)
<223> This region defines exon 8

<220>
<221> Intron
<222> (395)..(720)

<400> 13

agcagaacag ggggacggaa gtacatacac gttgtacagg tacgatcccc aaagggccac	60
cagggcagcc cgcagaggca cttgggccag agcctcctgt cttccccca gaagatgccg	120
caatgtcaca ccaccagctg actggggcta aaatacagtc aggattcaag gccagtccc	180
caagccatga ctgacccatg ttcccccaga ctgtcgtaacc ttagcaaagc catcctgact	240
ctatgttttgc tcaccaggaa acgcccagag gtcgtgacca taaccagatg gaaggctcca	300
gtggtatggg aaggcactta caacagacgt cttagataat tattatgcca aacagaaaat	360
taccgtggc ttgacggttt ttgctgtcggtt aaggttaggtt ttgctaataa aactggcctt	420
gagttttcc cttccacta tcagaggatg ggtgaggggc ccctgggttt acagaggctg	480
ttcatgtcat gtctgaatta gtggagagga gaatgggtgc acaggccat tttagactcc	540
cttctgctga ggtccccaaa ggctaagaat aaaactagtc agagggtcaa ctcttccca	600
cctcagggtg aggggcttgg gttgcaggga agaaaaatctg ctataccac tgcacccaaa	660
gtcgacagta cacccacagc cacctccacc ctgacaccttca cggccctctg tggaaattcc	720

<210> 14
<211> 2964

<212> DNA

<213> Sus scrofa

<220>

<221> Intron

<222> (1) .. (318)

<220>

<221> misc_feature

<222> (319) .. (2904)

<223> This region defines exon 9

<220>

<221> terminator

<222> (1010) .. (1012)

<223> This is the translation termination signal

<220>

<221> 3'UTR

<222> (1012) .. (2964)

<220>

<221> terminator

<222> (2858) .. (2863)

<223> This is one transcription termination signal

<220>

<221> terminator

<222> (2883) .. (2888)

<223> This is one transcription termination signal

<220>

<221> polyA_signal

<222> (2904) .. (2904)

<400> 14

tgcaatgcc agagcagctg aaaacacatg ttctctctgc ctggttggct tccaagagtg 60
agagaggaag gagcagggtc gagcatgcc agccaccctg ccagaatcac cagtcaggta 120
agccactcca cctccccaaa gctgaatgac tgaatggtgg agagtagctg ggaatgttac 180
agcaacagac gtctctcatc caggatgggg aaaaatcatt ccttcctaa actgcaaaat 240
acagactaga tgataatagc atattgtctc ctctagaaat cccagaggtt acatttaccc 300
cattcttctt tatttcagat acattgagca ttacttggag gagttcttaa tatctgcaaa 360
tacatacttc atggttggcc acaaagtcat ctttacatc atggtggatg atatctccag 420
gatgccttg atagagctgg gtcctctgct tcctttaaa gtgttgaga tcaagtccga 480
gaagaggtgg caagacatca gcatgatgctg catgaagacc atcggggagc acatcctggc 540
ccacatccag cacgaggtgg acttccttctt ctgcatggac gtggatcagg tcttccaaaa 600
caacttggg gtggagaccc tggccagtc ggtggctcag ctacaggcct ggtggtacaa 660
ggcacatcct gacgagttca cctacgagag gcggaaaggag tccgcagcct acattccgtt 720
tggccagggg gattttatt accacgcagc cattttggg ggaacaccca ctcaggttct 780
aaacatcact caggagtgtc tcaaggaaat cctccaggac aaggaaaatg acatagaagc 840
cgagtggcat gatgaaagcc atctaaacaa gtatttcctt ctcaacaaac ccactaaaat 900
cttatccccca gaatactgct gggatttatca tataggcatg tctgtggata ttaggattgt 960
caagatagt tggcagaaaa aagagtataa tttggtaga aataacatct gactttaaat 1020
tgtgccagca gtttctgaa tttgaaagag tattactctg gctacttctc cagagaagta 1080

30

gcacctaatt ttaacttttta aaaaaatact aacaaaatac caacacagta agtacatatt 1140
attcttcctt gcaactttga gccttgtcaa atggggaaat gactctgtgg taatcagatg 1200
taaattccca atgatttctt atctgttctg gggtgagggg gtatatacta ttaactgaac 1260
caaaaaaaaaa attgtcatag gcaaagaaaa agtcagagac actctacatg tcatactgga 1320
gaaaagtatg caaaggaaag tgtttgcaa caaaataaga ttgggagggg tcgtcctctt 1380
gatttagcg tcitcctgtc tctgctaagt ctaaagcaac agagttgctt tgcagcagga 1440
gatcagagtc taccttagca atcctcagat gattcaaca gcagaggact tcaggttatt 1500
tgaagtccat gtcctttcg catcagggtt ttgttggct tctgcgcagg atactgatca 1560
agattcccaa tgtgaatgtt ggagttacag ggaatccgaa tgaaccaatg ggagctcagc 1620
acgaaataaa agcacagctt ctaagtaagt ttgccatgaa gtagcgaaga cagattggaa 1680
agagaggggg ctgatcactg tggggcaatg ccatttctaa gagacacagg gcatggagtt 1740
ggcatgtaca tacagcttgg atccaggcac tgaatggag gcaatgagag tggctccagc 1800
ctcctaacc atatgacaac tagagcagca ctgtcttaga agatgcttct tgcttggcc 1860
aagtcatatt cagtctgcca gactctggaa ctgtgtcta caaatccttg ctcagaggaa 1920
gtggatgatg tcagagtggc cagaggccta cattgggttg aagtgacttc ctagaccttg 1980
gcttcatgac aatcaggcat cagcaagccc tgctgccacc tgctctaact ctcagagtcc 2040
ctcagcccat catgggcaac ttgagagcca ccgtcaagga gtggactaga ggaaaagcct 2100
gcttatcagg gaacctctca tttccctgc cccagctgca ctactgaagt gtaactgccg 2160
gacatgtta ataaagtggt taattgattt tatatcaaag tagagaggat ggcaatggga 2220

31

gaccaggc tcatgactaa acagctttc aatcccttc tctaagaaaa gctatgagat 2280
cttacatgt aattaaagtt aagcagttt ggttaaagga agttaggagg caatattac 2340
atctgcaggt atgtgatata ctttgctt gttccagtt taggtcattt gtgtccattt 2400
tcaaatgatt tacttgaaga gccattgcac tgacttgatg ttcagcacga tgggcttctt 2460
tgataaaatg aaacctacat ttctctact gttccctgg gcctcctact ctcaattct 2520
tgctaaaaat ttttgcacc cagcaaaata actcaacaaa ataacccaac aaaataactc 2580
aacaaaaatc ctggagaagt agtcttgtaa aagaaaaagg aaatcacaag tcaatttagga 2640
ctcttggttc tctataacgc aagtttatgg aatccattct ggagtgcaga gacttcattgg 2700
tgcaagttcc aaactacaga aatgattcgt tctcaaagat taaagaaaag gactgatatt 2760
tcctttgaa ggaatcttga tttttaaaaa aaaaatcatt taaattttaa tttcaaattgg 2820
acaaattcaa gatcttatta atagttcaat attaaaaaat aaaaattcct gattttaaaat 2880
taaataaatt attttctcag tatattctgg tctggtcatg gattgtggct tttttcccaa 2940
agatgttcag aactgtcatt taca 2964

<210> 15

<211> 1500

<212> DNA

<213> Sus scrofa

<220>

<221> misc_feature

<222> (1)..(1500)

<223> genomic sequence between exons 2 and 4

<400> 15

ggatccttaa gccactgagc aaggccaggg atgaaaccca caacctcatg ttcccttagtc 60
agattcgtaa accacagagc cacgacggga actcccacac attatttatt gacggccttc 120
tctgctctct gtggggact gggaaattcag gggtgatcaa gaagtcatcc ctccctgcct 180
caggaagctc aaaccactca ttatttattt acggccttct catgctctct gtggggact 240
ggaaattcag gggtgacgaa gaagtcatcc ctccctgcct cacgaagctc aaacaagcag 300
gtagaggagg cagagcaaaa tgcaggtctt atccggtgag ccgactcccc gggcgatgtg 360
tacagcaaag gaatagaggg atggggccg gaggagagaa aagggtttca gccgtggtca 420
gggtgggggt gggaaagtggc ttcacaaagg cagtgacatt ggctccagg tgtccactct 480
tctgtctctg ctaccttctg gtcctctcct tgtggccct cctctatcct acctctaaag 540
cttcagccccca gcaccccttctt ctctctctgc attctctcct gggtaatcaa 600
attcgttccc ttacacgtcag atccggtatac ttcccttggtc catgaacaac ttctccgatt 660
gcacggtctg cctacatctc tctgatgaac tttagacttg aatgtccact tgtctccctg 720
tcccctttta ggtattcgca cactccccga cattcacacg tccaaaagg aattcatgat 780
tattatcctc caagcctgtt cctcctccag cccatctgag aaaatactac aaccccccctg 840
cttaagcaga aatcttgggt ctccctgtc tcatctctga taacaaaatt accaaccacg 900
tccttatcaat tctctctcca aagtatataat atatatattt ttttaattt ttcccgctg 960
tacagcatgg ggatcaagtt attttacat gtatatttc ccccccacctt ttgttccgtt 1020
gcaatatgag tatctagaca tagttctcaa tgctactcag caggatctcc ttgtaaatct 1080
aagttgtatc tgataacccc aagctccccga tccctccac tccctccctc tcctgtcggg 1140

cagccacaag tctattctcc aagtccatga ttttctttc tgtgggatg gtcatttg 1200
ctggatatta gattccagtt ataagtgata tcataatggta tttgtcaaag tatataat 1260
atttttcttt gtcttttgt cttttgtctt tttttgttg ttgttgtgt tgcgttgt 1320
gttggttgcta ttacttggc cgctcccgcg gcataatggag gttcccaggc taggagtt 1380
atcggagctg tagccaccgg cctacgccag agccacagca acgcgggatt cgagccgcgt 1440
cgccaaccta cacacagctc acggcaacgc tggattctta acccactgag caagggcagg 1500

<210> 16
<211> 500
<212> DNA
<213> Sus scrofa

<220>
<221> misc_feature
<222> (1)..(500)
<223> genomic sequence about 4-5 kbp downstream from porcine exon 4.

<400> 16
ggtaccatg aaaagccaa caacacaggc tagaaggagg atgtcagaga gagagagcaa 60
aggaacgtga gagttcaggg agggcaaggt tatgtttggc ttggagatgg atctatgttt 120
tgcatttatt ttttggggg ggggtcttt tgctacttct tgggctgctc ccgaggcata 180
tggaggttcc caggcttaggg gtctaattgg agccgcagcc accagectat gccagagcca 240
cagcaacgca ggatctgagc cacgtctgca accttcacca cagctcacgg caacgcccaga 300
tcgttaaccc actgagcaag ggcaggacc gaacctgcaa cctcatggtt cctagtcaga 360

34

ttcgttaagc actgcgccac gacggaaact ccctcattt aaaaatattt ttgagcacct 420

actgtatgcc aggcatgtt ctaggttcat accaaagaag gctcaaaaag atggcatccg 480

aactgttgcc cttgaaagga 500

<210> 17

<211> 1520

<212> DNA

<213> Mus musculus

<220>

<221> Intron

<222> (1)..(1130)

<220>

<221> promoter

<222> (381)..(1321)

<223> Fragments and derivatives have promoter activity.

<220>

<221> 5'UTR

<222> (1131)..(1320)

<223> untranslated exon 1

<220>

<221> Intron

<222> (1321)..(1520)

<400> 17

tcccaatgca tctttccca gtgggctctt ggattcatgc tgccatatga tctgctgata 60

ccatgcttca gtaccaagtt gattcttgc tcttgtcctg atgctgaaga cctaaaatga 120

tgaaaatgga aaaagaatga agaataagta tacacacacc cggcctgctt ttgcggatca 180

ggtgtggtccc gccgggcgtc tgacactgaa gccacgcggg ggcttcagtg gggaggaggt 240
gtgggcgagc gcgagcgccg ctattccggc ccagccctac ctcggtcctt gctttgtcc 300
tggtcactcg atcatttcct ctgtatccac ttctgaactc taggctctgt cccaccctga 360
acagtgtcgc tgcatactgtt tgcttactgg ggtctccgc cacctccct cgctatccga 420
atagctgata ttcagggcag cacagggcag ggcagggcag ggcagggcga gtagggcaga 480
tcagatcctg ggaccaccgg tactaaccag tgagtgtaga aagcaggagg tgtctttcc 540
tactgttagtt aggacagggc gggttggctc ttcttatgga caagatggaa aaggggtgca 600
ggtagggca aagtgagaga cactcgaatt tgagagacag acagactcct aacagtgaag 660
gaaggaccaa gccaaaatca agcctggca aagtctcagg cactaacttt gctgtgtgg 720
gtgatggag gtaatctcg tacaacttt caaaccacct cgttccact gcaaggagac 780
accatcaagt gttgaagat ggcagggaa cctctcaaca aaacacacac acaaacgttt 840
tattatTTA tatttatttt gcatgcaaag tactgtgtt cattatggca tttcataca 900
tatgcgattg cacaactct tggaaatcat ccaagaaaca gcaaagcggg aaataatgtt 960
gtgggggggg ggcgcggagg agagagaaca gagactggag agagtgtgt ctccttgct 1020
gcggggggcca ggaagaggct aggagggcgg ggtatgtcaac gccactagct ctcctcag 1080
gaaggacccc agggactctt attttgttag tttgcttgt ctggccact atcgccccca 1140
gaacagatct gactgcctct ttcattcgcc cggaggtaga taggtgtgtc ttaggaggct 1200
ggagattctg ggtggagccc tagccctgcc tttcttagc tggctgacac cttcccttg 1260
agactcttct tggaaatgaga agtaccgatt ctgctgaaga cctcgcgctc tcaggctctg 1320

ggtaggcaaa ggcgaggggg ctcgccatgg ctcgggttgc ccagggattg gggcatcagg 1380

actacggag tctctgcctt ttgatagtgc ttcccttacag ttattttgg gagtagttgc 1440

ttcttcctga tggagccgc gtgcgggtcc aagctatctt ttgcaagtaa caggtgtctg 1500

nnnnnnnnnn nnnnnnnnnn 1520

<210> 18

<211> 1207

<212> DNA

<213> Mus musculus

<220>

<221> Intron

<222> (1)..(653)

<220>

<221> 5'UTR

<222> (654)..(773)

<223> untranslated exon 2

<220>

<221> Intron

<222> (774)..(1207)

<400> 18

agccctaggt tgcgttggc tacacagtga gttcataggc tgctaggat cctatctcaa 60

aaaggaaaac aaacaaacaa acaaagggtg ggcagggtt agccttgtcc ctcaggagca 120

ggtatggtt tctgaggctg tcccaagtgc atatggtaaa ggcttctcta tggagattt 180

caccatttc taaagtgcag tttccacat aactgtgtgg cttccagagc caggctgtgg 240

37

aggaagagct tatctcagaa ccacatttg gcgtcccatc aaagtgcctt gtccgctaac 300
ctgcctctgc cccaggctgt gtcatacgca tctccgggaa ggcatacatt gagaatgagt 360
gcatacaca gggctcccag tttcccttg ggactgggtg atgtggaggg tggtggcctc 420
atcgcttgtg actcctggca tggttggtc ctgcagttt tcctctgggt gaggaagtca 480
gaggaccaac ccagagccct gattctgcct tgctgcgtag acctgaatca acagccctga 540
taaacagccc attccccggg gctgagggaa caaagcctgt ggctgctgcc gagggatctg 600
tctgccacc accaaaaatcc tcttcctgaa acagctgtt attattttga caggagttgg 660
aaccctgtac cttccttcc tctgctgagc cctgcctcct taggcaggcc agagctcgac 720
agaagctcggtt ttgcggcgtt gtttgcgtt gaggaaacac agctgacgat gaggtatgtt 780
taaaggattt gtgtctccca gccttgggtc actgcgagct actgttaggt caccaaatgg 840
ttccacctga gggaggaccc ttgctcttcc ccgaagcttt ccttggtccc ttctgtgatt 900
tgttgtccctt tccctttgtt ttctgaaaca ggggctggtg gaatgctggc tggggacttt 960
ggtattctgc ttctcttggc agccccccggg gctatgccag tcaaggctgc agcctggagt 1020
tctctgtgtg gggtttgggt tggcggggct gagtcttggg cagggcgcgg tgggagggtg 1080
ctgagtccttc tctgctctgg gctgtctcgat acatgtccgt tggctggctg ttccctggag 1140
gtatcacttg agattgattt cattccacat gacactgctc ccagggacag cccggcactc 1200
nnnnnnnnn 1207

<210> 19

<211> 900

<212> DNA

<213> Mus musculus

<220>

<221> Intron

<222> (1)..(336)

<220>

<221> misc_feature

<222> (337)..(517)

<223> untranslated exon 3

<220>

<221> Intron

<222> (518)..(900)

<400> 19

ttccggcatc tttaagatct ttagtaccc aagtcaactt cagcttcaca gcttcttgtt 60

tcaatgtctg ggatccacac ctgatcttct gggatcctcc aaagggcttg ggtcacttct 120

ccatctctgc cctctgttagt actctaggct ctatgtact ccactccact gctgctgctg 180

ttcttggtga tcatcctatg gtactggcaa gtaggtgaaa gaagaagagt gaatattcct 240

tcacccaatg tccttatgta ggcctccagc agaagggtgt gctcagatta aagggtgtcta 300

cccccatgcc tggatctaaa acttgctttt ttccaggctg actttgaact caagagatct 360

gcttacccca gtctcctgga attaaaggcc tgtactacat ttgcctggac ctaagatttt 420

catgatcaact atgcttcaag atctccatgt caacaagatc tccatgtcaa gatccaagtc 480

agaaacaagt ctccatcct caagatctgg atcacagggt tgcccttctg tttctggatt 540

atagttcatc ccagatgttag tcaagttgac cactaggaat agccatcaca agcccgttgt 600

ggaggctgcc ccctgcccc cgccccgcgc gcccctgagg ctctcacccc tttcttgtgg 660

cagctttgtt cttcatctcc agtgtacaac tgcattccc actctgcata ttgccttcct 720

gaacaatcac cgccccaaag ttcttctcag tcctttgtta tcctcttcctt tttcttcac 780

aatcttatgc agaatttaaa aaataacctga ctcccttcagt agttccagtt gtttgctggc 840

ttgtgggggg ttagtgggg tgaaccagggt ggggctgaaa agtgggtgca nnnnnnnnnn 900

<210> 20

<211> 1020

<212> DNA

<213> Mus musculus

<220>

<221> Intron

<222> (1)..(479)

<220>

<221> misc_feature

<222> (480)..(568)

<220>

<221> 5'UTR

<222> (480)..(489)

<220>

<221> misc_feature

<222> (490)..(492)

<223> This "atg" is the translation initiation codon

<220>

<221> Intron

<222> (569)..(1020)

<400> 20

40

tctcttcca atgcccacat ggatgggctt cagcatcatt tcagatcatg aagcctcatt 60
aactgtgctg gcctaattgg ccatgactag tttgtgtgct tgagggatag ggggagggga 120
gacacttgtc gctgagttag ttacaaatgt atccctgttag gaaggatgtg ggcagatgcc 180
tttcattatac tttaactgcat caaacatttt atgggtatga gtgtttgcc tgcaagtatg 240
tatatgtacc acttgtatata tgggacccca tggaggccag aagagcatca ggtcctgtga 300
aaccagagtt atggacacct gtgagctgca aatgtggatg ctgggaactg aatcgagcag 360
gtgttcatt gaggtgttcc aaccacacag ctgtttctcc agccccagaa gccatctctc 420
attccagatt tagtttattt aatctatttc cccctctttt tttctccctg cctctacagg 480
agaaaataat gaatgtcaag ggaaaagtaa tcctgttcat gctgattgtc tcaaccgtgg 540
ttgtcgtgtt ttggaatat gtcaacaggt aattatgaag ccagctagaa aggctgcttt 600
cattccctgt gactggtgcc agctgagtga ccaatcagtc tgaacataag ggacggagcc 660
gtgagcagga gtccagtctt cctgtgttcc tgagccccag atggccatta aaactgtaga 720
ccatccaagt cacttctgcc ttagtaatta tcctcttca tgccgtgctc ctcaaaccctc 780
gaatttctgt aagctagatg gagagagaaa gtacattaag ccaaaaccac catctcaagt 840
aatttgtata agcagatccc agaagattca ggccaggcag ggtagtgcat gtatggagtc 900
cttgtgcttgc caaggcagag gcaggagcat catacaaattt gaagaccaag cttgtcttca 960
tagtgacttc caggccagct gtgccttac aaggagaccc nnnnnnnnnn nnnnnnnnnn 1020

<210> 21

<211> 1020

<212> DNA

<213> Mus musculus

<220>

<221> Intron

<222> (1)..(584)

<220>

<221> misc_feature

<222> (585)..(620)

<223> exon 5

<220>

<221> Intron

<222> (621)..(1020)

<400> 21

tggccacact agcttttac cagttttcc caggcaaatt ccttagccag gatgtatgtt 60

gctgtatgtt gccttctctg ttacattgtt tattttcat gagccccagc actcggtgtg 120

taggacttgc cttagcacgtg taagactgtt gagagctgtt gccctaaagt agttgttagct 180

ggccttagcct tctggtaaaa gcaacaccca tgggggctgc tcagaagaag ggatctgagc 240

tgaatgtggc ggctatttcc tgtgggaag aatcctcagc ctgaggtggc tggccgtggc 300

/

gcttccacct tccccgcctt cctcattgcc cagttctgg gactgtggt gaagaggacc 360

ttcctgtcat gtaacaaaca gctgggtgac tttaaaagag agaaagaggg aaaaaaatcc 420

cccaaataaa aacaagaatt gagagtgtt gggtgcccac ttctgttcct cagtgatgct 480

tgtggaaatc ccctgagaac ccaaacgctt aaggaaaacc actgcagtga agctttctg 540

agaattaaaa gtatatgacg tttctatttc ttatttgtcc ttagcccaga cggcttttc 600

ttgtggatat atcacacaaa gtaagtgttc tgaattctgt gtatctattg gatgtctgga 660

tcacttgatt tttttttt agcccctaaa gttgattcc tctcttcaag ccagccaatg 720
tagtgctcg gccacagtaa aggaggaga gagggcagga cagggaggag gattgctagg 780
gccctgggtt cagggctgca actctgctag tccccaaact ggtctttgt aataatgtat 840
gagtttgct ctgggttcg ctcagggac tctcctcaaa tattgtcatg gggaccattt 900
ttggttgacg tagggaaaga gcccaggaa ctgcattgtg tagtgtgtac cctcagtgt 960
gctgtgaggc actgagggag gacttacgtt cagttccagt nnnnnnnnnn nnnnnnnnnn 1020

<210> 22

<211> 1020

<212> DNA

<213> Mus musculus

<220>

<221> Intron

<222> (1)..(595)

<220>

<221> misc_feature

<222> (596)..(661)

<223> exon 6

<220>

<221> Intron

<222> (662)..(1020)

<400> 22

cagatttcct gagtttcat tgattggca atggatttt tttctcagat taatcttat 60

aatacatgca tgtatacaga cacacacaga cacacacgca tgcagtcatt ctcggaaagg 120

43

tgcttttct tatttaata ttacccctcg ttacagccgc tttatgtca ccaggctctt 180
gcatatctgc tgtctcattg gtcattacag atcccttcgt ggacggatta ttattgatta 240
ccctttcag agaagaacgt ggcagtttag acagtgtgag tgtatgcaa agtcaactcca 300
ctagcaggag gagatcgtga ccacaggctc tcagatctgc agggtctcca ccattctgat 360
ttccctgccc cttatcccttc aggggtccca gggatgagca gagtgctcag ggctgcccag 420
aagggcgcag ctgaggcccc tcaagtccac tctctgcctt tagctcagct gcctttgcg 480
tgtccatgtt tcatgagctg catcttgacg ttcaactttt cttagtgc tac ccgaccctta 540
aagttcagga ccgcctcgat ttctagatgt gtttatattc ttttcattt cctagaattc 600
cagaggttgg tgagaacaga tggcagaagg actggtggtt cccaaagctgg ttaaaaaatg 660
ggtaaggat caggaatggg tcctaagtcc ctgaaaccca cagaggaccc atggcctcct 720
ccctcccttc ttctggctca ctggactcac tcatggagtc tccccattgc tgggtttgtt 780
tttgggttg ttagcttcta ttgttattgt gaggggtggg gagtgtgtt gtgtgtatga 840
cgtgtgtatg attgcagctg tgtgtacacc atagtactca tcggaggtca gaaggcactt 900
tcaggaggca attctgcctt tccagtaacgg gttccagtg gtgatcacca gactcagatg 960
ctcaggcttt caggacaagc agtttacag gatgagccat nnnnnnnnnn nnnnnnnnnn 1020

<210> 23

<211> 912

<212> DNA

<213> *Mus musculus*

<220>

<221> Intron

<222> (1)..(389)

<220>

<221> misc_feature

<222> (390)..(491)

<223> exon 7

<220>

<221> Intron

<222> (492)..(912)

<400> 23

ccttcttccc actccctcct ccccccccttc ctgtctgcct tcctccttcc tctgcccttt 60

cttctccagt taagggtgaa gttcaggctg aagtggaaat ttcagaatacg acacagaaca 120

gaaatgtccc ttggagttact gttctgaaac atctcaccga cttctgaaat aactgagggt 180

tacagggtca ctggaacctc agccctgac ccacatggtg gccagagagg caaatgctgt 240

accttttatac agaagtgtgt agggatcaag gggtcagtgc cctgagtcct ccagtccacc 300

cagtgggtgtg agtgatgcct tcttccctt gagacacgag tcatggaagc cactgtcct 360

taccaacttt gtcctacctt tgtccacagg acccacagtt atcaagaaga caacgtagaa 420

ggacggagag aaaagggttag aaatggagat cgcattgaag agcctcagct atgggactgg 480

ttcaatccaa agtaaggacg gacaggagat tggggtgggg ggtgctgagt ggggtctga 540

ggagatgctg aggggagtgc tgaggggttg ccggcaggag ggggtgctgg caggagaggg 600

tgctggcagg aggggggtgct ggcaggaggg gatgctggca ggagggggtg ctggcaggag 660

ggatgctgg caggaggggg tgctggcagg aggggatgct ggcaggagtg gtagacctt 720

cctcaatggg ctttggctaa gaaactaaga tctgggtgct ttgaaccaga ctgaacactg 780

tggtaattgc agcaggaaat ggccagtggt aggttaaaca taaacactgg gtgttaagga 840

ctttacaggc cacataggat gctgctgaga aaatgacaag gtctaagggt gagccaagaa 900

nnnnnnnnnn nn 912

<210> 24

<211> 608

<212> DNA

<213> Mus musculus

<220>

<221> Intron

<222> (1)..(221)

<220>

<221> misc_feature

<222> (222)..(359)

<223> exon 8

<220>

<221> Intron

<222> (360)..(608)

<400> 24

catccaggac ctactatctt tgtacttcac tctgtgtcaa gagttggagg taccctacgc 60

atttgtgcct ggcccttgcc aagactccac cccttctgtta cttcctgtct ttcatgcagg 120

caagatttcag tgacagtcac tggcctccct tccttggcca gtctctcacc acacacctgt 180

gtaatgcttc tgactcggtg ttgcattgtt ctctcacca ggaaccgccc ggatgttttg 240

acagtgaccc cgtggaaggc gccgattgtg tgggaaggca cttatgacac agctctgctg 300

46

gaaaagtact acgccacaca gaaaactcact gtggggctga cagtgttgc tgtggaaag 360
taagcaccac tgacaaaactc acccttgatg atttgttctt gttctagcat caaaggatt 420
gtgtgggct ccagggcccc acaaaggctg gaatttgaca gtagacttcc cccttcttc 480
ttataatggc tgagaaaaaa caatgatagt aggtgatgag gtatttctct gccagtgagt 540
gagccaatcc aagccagagt agattgtatt aaatacaggt ttattggaa gctgctctca 600
nnnnnnnnn 608

<210> 25
<211> 3240
<212> DNA
<213> Mus musculus

<220>
<221> Intron
<222> (1)..(369)

<220>
<221> misc_feature
<222> (370)..(3010)
<223> exon 9

<220>
<221> terminator
<222> (1088)..(1090)

<220>
<221> 3'UTR
<222> (1091)..(3010)

<220>
<221> Intron

<222> (3011)..(3240)

<400> 25

ttagcagata cactggcctc ttctggatat tcaagagcta gctccttctc tgacagccag 60
cttctcaatc agagaacaga gccttagcat gaaccttact gcaacgcaga gtagttgaga 120
acaccgagct ctcagtgtgg caggcatcga agagcacgcg gtcggggctg tgcatcccc 180
gttgcttaa caaagctggc agtgagataa gtcatgccac tttcccaag gacacaatga 240
ccagctagtg tcgagtggta tgtggagaag ccatcccctc ctaacataca atacagatca 300
tctactgtaa tgttaagtat ggtattacat gtatatatgt acccatatat aagtgtgata 360
gtccgtggtg gtcaatgta gccctctcta tttcaggtac attgagcatt acttagaaga 420
ctttctggag tctgctgaca tgtacttcat gggtggccat cgggtcatat tttacgtcat 480
gatagatgac acctcccgga tgcctgtcgt gcacctgaac cctctacatt ccttacaagt 540
ctttgagatc aggtctgaga agaggtggca ggatatcagc atgatgcgca tgaagaccat 600
tggggagcac atcctggccc acatccagca cgaggtcgac ttcctttct gcatggacgt 660
ggatcaagtc tttcaagaca acttcgggtt ggaaactctg ggccagctgg tagcacagct 720
ccaggcctgg tggtacaagg ccagtcccgga gaagttcacc tatgagagggc gggactgtc 780
ggccgcgtac attccattcg gagagggggta ttttactac cacgcggcca ttttggagg 840
aacgcctact cacattctca acctcaccag ggagtgcattt aaggggatcc tccaggacaa 900
gaaacatgac atagaagccc agtggcatga tgagagccac ctcaacaaat acttccttt 960
caacaaaccc actaaaatcc tatctccaga gtattgctgg gactatcaga taggcctgcc 1020
ttcagatatt aaaagtgtca aggtagcttgcagacaaaa gagtataatt tggtagaaa 1080

taatgtctga cttcaaattt gatggaaac ttgacactat tactctggct aattcctcaa 1140
 acaagtagca acacttgatt tcaactttt aaagaaaaca tcaaaaccaa aacccactac 1200
 catggcaaacc agatgatttc tcctgacacc ttgagcctgt aatatgtgag aaagagtcta 1260
 tggcaagtaa tcaggtataa attctcaatg atttcttata tattctgggt ctggggaaaa 1320
 cttgattcta gaaatcaaaa ttaatttgc aaaggaaaag cagatgccgg aaacttcttc 1380
 ccagtctgtc atacaattca ccactggcca ggtgctgaga gaagcattag ggaacagtgt 1440
 gggtttgtc agagttggac ggctccatcc cttggcttc attatcttcc tcctcatgga 1500
 gattctaaag caacccagag aggcttgca gccagagacc tttataagg atgccaatgt 1560
 gaccatcagt ctgtaaaagc ttagggctcc aggagcgctg gcagtccagg ccccactagg 1620
 ctattgttc tgtcctgggc ataaaggagg cagagagtgc caataggtac tttggggca 1680
 catgttcaga gtccaggaaa aatcaagggt gaccacttag agggacatag gacttgggt 1740
 tggtgattga actgagttac aaacacagac agcttcttc aggatgacta acagcaggaa 1800
 ttgaatggaa agtgtgttca ttttggggcc cccaaattgt attcatgctg ttagctttgt 1860
 gtgttggcc ctgtggagag ggtgtgactg tatcagggaa ggagagtacc tcagcggact 1920
 gaggaccagc accctattat atcagaagac aatctctcat catcaggtcc tacctacaac 1980
 ctgctctgaa cctccgagtt cctcagccca tcgtgttcca gtgtggggc ctgtatggag 2040
 caggtgactg aagacaaagc cccctgtcac atgacctcat ttccccctgct ctgtactat 2100
 gcaagtgtga cagccagccca gccagatgta ctggacaaca taggaaccga ctttatggca 2160
 atgggagccg cagtcactac aacggagctg ctgaaggttc tggtccccgc tctgagagcc 2220

tgcaggagcc cctgtatagg tggttctcaa cctatgggtc gcgaccctt tgggaagtgt 2280
taaatgaccc tttcacaggt gtcccctaag acggtaaaa aacatagata ttccactct 2340
gactggtaac agtagcagaa ttacagttat gaaatagcaa gggaaataat tctggggttc 2400
gtgtcatcca taccatgagg agctacatta ggtcacatca ttaggaaagt tgagaagcat 2460
agctctactt gggatttaa gcaaattatg caaagggggt tgtcgtctg tttctgtgt 2520
atgcataatat ttatattttg ctgtcttcc agtttaggtc aatctgttcc ttcccttaag 2580
cagtttattt aaaaggccat tgcaccatct tggtaacag catgagggtt ttcaataaaa 2640
aataggatct taccttgtc cacaggctc tacctcttac tttcaattt gtaacaaaaa 2700
aggtcgacaca cccagaggca acaaaaccca cagaattcct gaaccaatgg gagatgccaa 2760
tggaaagcaga gcttgcacat ctgctaaaaa ttctgcctct ctgtcaactgt gctggatccg 2820
tctaaagtgg gacagttcaa tggctgaaa gttcaaaaa ggctggggaa tttgagggga 2880
tttttttta aaataaaatt gatccaagtt taaatctcta atgagtaagc ttaggatttt 2940
attaaaggta atttttagac attcttcaaa ataagaattc ttgtttataa ttgaataaaat 3000
tattttctca gtatattttg gtctggatg gattatgcgt tgtatcctga agatgttcag 3060
aagtgtcagt tgtgattgtc cataatcata aaggattta cgataccttgc aatgagcttc 3120
acaaagacaa gattacaaag aacaggctt attctcaaattataaagtgt gctctctctc 3180
aatctctctc tctctctctc nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 3240

50

<212> DNA

<213> Mus musculus

<400> 26

atcggcccca gaacagatct gactgcctct ttcattcgcc cggaggtaga taggtgtgtc 60

ttaggaggct ggagattctg ggtggagccc tagccctgcc ttttcttagc tggctgacac 120

cttcccttgt agactcttct tggaatgaga agtaccgatt ctgctgaaga cctcgcgctc 180

tcaggctctg ggagttggaa ccctgtacct tccttcctc tgctgagccc tgccctccta 240

ggcaggccag agctcgacag aagctcggtt gccttgctgt ttgctttgga gggAACACAG 300

ctgacgatga ggctgacttt gaactcaaga gatctgctta ccccaagtctc ctggaattaa 360

aggcctgtac tacatttgcc tggacctaag atttcatga tcactatgct tcaagatctc 420

catgtcaaca agatctccat gtcaagatcc aagtcagaaa caagtcttcc atcctcaaga 480

tctggatcac aggagaaaaat aatgaatgtc aaggggaaaag taatcctgtt gatgctgatt 540

gtctcaaccg tgggtgtcgt gttttggaa tatgtcaaca gcccagacgg ctctttcttg 600

tggatatatc acacaaaaat tccagagggtt ggtgagaaca gatggcagaa ggactgggtgg 660

ttcccaagct ggttaaaaaa tgggaccac agttatcaag aagacaacgt agaaggacgg 720

agagaaaaagg gtagaaatgg agatcgcat gaagagcctc agctatggaa ctgggtcaat 780

ccaaagaacc gcccgatgt tttgacagtg accccgtgaa aggccggat tgggtggaa 840

ggcacttatg acacagctct gctggaaaag tactacgcca cacagaaact cactgtgggg 900

ctgacagtgt ttgctgtggg aaagtacatt gagcattact tagaagactt tctggagtct 960

gctgacatgt acttcatggt tggccatcggt gtcataatgtt acgtcatgtat agatgacacc 1020

51

tcccgatgc ctgtcgtgca cctgaaccct ctacattcct tacaagtctt tgagatcagg 1080
tctgagaaga ggtggcagga tatcagcatg atgcgcata agaccattgg ggagcacatc 1140
ctggcccaca tccagcacga ggtcgacttc ctcttctgca tggacgtgga tcaagtcttt 1200
caagacaact tcggggtgga aactctggc cagctggtag cacagctcca ggcctggtgg 1260
tacaaggcca gtcccggagaa gttcacctat gagaggcggg aactgtcggc cgctgtacatt 1320
ccattcggag agggggattt ttactaccac gcggccattt ttggaggaac gcctactcac 1380
attctcaacc tcaccaggga gtgcttaag gggatcctcc aggacaagaa acatgacata 1440
gaagccccagt ggcatgatga gagccacctc aacaaatact tcctttcaa caaaccact 1500
aaaatcctat ctccagagta ttgctggac tatcagatag gcctgccttc agatattaaa 1560
agtgtcaagg tagcttggca gacaaaagag tataatttg tttagaaataa tgtctgactt 1620
caaatttgta tgaaaacttg acactattac tctggtaat tcctcaaaca agtagcaaca 1680
cttgatttca actttaaaaa gaaacaatca aaacccaaac ccactaccat ggcaaacaga 1740
tgatttctcc tgacaccttg agcctgtaat atgtgagaaa gagtctatgg caagtaatca 1800
ggtataaatt ctcaatgatt tcttatatat tctgggtctt gggaaaactt gattctagaa 1860
atcaaaatta atttgacaaa ggaaaagcag atgccggaaa cttctccca gtctgtcata 1920
caattcacca ctggccaggt gctgagagaa gcattaggaa acagtgtggg ttgtgtcaga 1980
gttggacggc tccatccctt tggcttcatt atcttcctcc tcatggagat tctaaagcaa 2040
cccagagagg cttgcagcc agagacctt aataaggatg ccaatgtgac catcagtctg 2100
taaaagctga tggctccagg agcgctggca gtccaggccc cactaggcta ttgtttctgt 2160

52

cctgggcata aaggaggcag agagtgc当地 taggtacttt ggtggcacat gttcagagtc 2220
caggaaaaat caagggtgac cacttagagg gacataggac ttggggttgg tgattgaact 2280
gagttacaaa cacagacagc tttcttcagg atgactaaca gcaggaattt aatggaaagt 2340
gtgttcattt tgaaaaatcc aaattgtatt catgctgtta gctttgtgtg ttgagccctg 2400
tggagaggggt gtgactgtat caggaaagga gagtacctca gcggactgag gaccagcacc 2460
ctattatatac agaagacaat ctctcatcat caggtcctac ctacaacctg ctctgaacct 2520
ccgagttcct cagccatcg tggccatgt tggtttttgt tatggagcag gtgactgaag 2580
acaaagcccc ctgtcacatg acctcatttc ccctgctcta gtactatgca agtgtgacag 2640
ccagccagcc agatgtactg gacaacatag gaaccgactt tatggcaatg ggagccgca 2700
tcactacaac ggagctgctg aaggttctgt tccccgtct gagagcctgc aggagccct 2760
gtataggtgg ttctcaacct atgggtcgac accccctttgg gaagtgttta atgaccctt 2820
cacaggtgtc ccctaagacg gttaaaaaac atagatattt ccactctgac tggtaacagt 2880
agcagaatta cagttatgaa atagcaaggg aaataattct ggggttcgtg tcataccatac 2940
catgaggagc tacatttagt cacatcatta gggaaaggta gaagcatagc tctacttggg 3000
tatataagca aattatgcaa aggggggtgt cgctctgtgt tctgtgtatg catatattt 3060
tatattgttttgcctt gtcttcagg ttaggtcaat ctgtttcttc cttaaggcag tttatattaa 3120
aggccattgc accatcttgg tgaacagcat gaggggtttc aataaaaaat aggtatctac 3180
ctttgtccac agggctctac ctcttacttt tcaattgtga aaaaaaaaaagg tcgcacaccc 3240
agaggcaaca aaacccacag aattcctgaa ccaatggag atgccaatgg aagcagagct 3300

tgcacatctg ctaaaaattc tgcctctctg tcactgtgct ggatccgtct aaagtgggac 3360
agttcaatgg tctgaaagtt tcaaaaaggc tggggattt gaggggattt tttttaaaa 3420
taaaattgtat ccaagttaa atctctaattg agtaagctta ggattttatt aaaggttaatt 3480
tttagacatt ctcaaaaata agaattcttg ttataattt aataaattat tttctca 3537

<210> 27
<211> 3135
<212> DNA
<213> Homo sapiens

<400> 27
gtggctgatc agagcgcgta gggcttcgcc ggggccccga gctggcgcg gtcctgctca 60
gcccagctca ccgcgcgcgc gcccctcgccg ccctcggcgc cctgggtctg cggatcagga 120
gaaaataatg aatgtcaaag gaaaagtaat tctgtcaatg ctgggtgtct caactgtgat 180
cattgtgttt tggaaattta tcaacacgac agaaggctct ttcttgcca tatatcactc 240
aaaaaaaccca gaagttgatg acagcagtgc tcagaagggc tggtggttc tgagctgggt 300
taacaatggg atccacaatt atcaacaagg ggaagaagac atagacaaag aaaaaggaag 360
agaggagacc aaaggaagga aaatgacaca acagagcttc ggctatggga ctggttaat 420
ccaaaatata atgatcatta cttggaggag ttcataacat ctgctaatacg gtacttcatt 480
gttggccaca aagtcatatt ttacatcatg gtggatgatg tctccaaatctt ggcgtttata 540
gagctgggtc ctctgcattc cttcaaaatg tttgaggtca agccagagaa gaggtggcaa 600
gacatcagca tgatgcgtat gaagatcact ggggagcaca tcttggccca catccaaacac 660
gaggtcgact tcctcttctg catggatgtg gaccaggctt tccaaatccca ttttgggtt 720

gagaccctag gccagtcagt ggctcagcta caggctggcg gtacaaggca gatccctatg 780
actttaccta ggagaggtgg aaagagtcag caggatacat tccatttgc caggggattt 840
ttattaccat gcagccattt ctggaggaac acccattcag gttctcaaca tcacccagga 900
gtgcttaag ggaatcctcc tggacaagaa aaatgacata gaagccaagt ggcatgatga 960
aagccaccta aacaagtatt tccttctcaa taaaccctct aaaatcttat ccctaaaata 1020
ctgctggat tatcatatag gcctgccttc agatattaaa actgtcaagt gatcgtggca 1080
gacaaaagag tataatttgg ttagaaataa tgtctgactt caaattgtgc cagtagattt 1140
ctgaatttaa gagagagaat attctggcta cttcctcaga aaagtaaacac ttaattttaa 1200
cttcaaaaaa tactaatgaa acaccaacag ggcaaaaaca taccattcct ccttgttaact 1260
tggggcttg taatgtggaa gaatgaatct agggcaatca gatataaatt cccagtgatt 1320
tcttatctat tctgggtttggggaaatac tatcaactga accaaaaata acttgtcata 1380
ggcagagata aagccagaaa cactctacac atgccagatg acatctggag aaaagggtgc 1440
taagggaagc gtttggcagc aagatatgtat tgtaagggtt tgcccttga gttcaatgtc 1500
tgcctatttc tgatgggtct aaagcaacat ggagttactg tgccagcagaa ctctcagtaa 1560
agacaccatt tgccttggca atcctcaaaa agttcaata gcagattgct tcagaccatc 1620
tgttagtccgt cctttctca tctggatgtt gttggcttc tgtgcgaaag attggtggag 1680
tgtcccagta gatatcatgg tggtgtgtga tcagagtccc aaggaacctg aatgagccaa 1740
ggtgcccagc atgaagtcaa aacaaagcct tgacatgagt ttgccatgaa atagcgaaga 1800
gagagtggaa gagaggagcc aatcactgtg gggcagtgcc accctgaggg cacttagggt 1860

atggggttgg tgcttaaata catcacagat ccaggtactg aatgggagga agtgtgggtg 1920
atttccaaatc tcattgaccc tatgttcagg gacttgaacg gaagatgttt ctgtgttgc 1980
ctaagtggta ttcaagtctac cagactctgc aacttgcac ttc当地atcct tggtaaagag 2040
atgtggatgg tgtcagagaa ggcaaaggcc tgcagtgat tgaagaggct tgcaaggcgt 2100
tctgtttcta ggatgtgggc ttcatcagaa gacactcggt caccacttag ctagtctaaa 2160
cctcagggtt cctcagccca tcataccccca acttggagga ctgacatcaa ggagtagact 2220
ggagaaacag ccctcccatc aagtaacctc ttgttctctc ctgctccatc tgcactata 2280
aagtgttaata attagacata ctggcaaaa tggctaattt atttggtaac agaagcatga 2340
gccataacaa tggaaagatct agtttatcatg actgaacacgc ttaacattca attcccttct 2400
ctaagagaag ctgtgaaatc ctacatatta tt当地aaagtta accaaatcaa tgtaaaggga 2460
gttaggagac agtgtgtacc tatgcacgta tatttatgtt ttgttgtgt tccagtctcg 2520
gtcatttgtt tccatattca agcaatttat tt当地agagcc attgcactag cctgatgtat 2580
actgcaatga gcttcttga taaaatgaaa ct当地aatttt tctcgaccat tt当地accgtgc 2640
ctcctacttc atttttgcc agaaaatctc acatccaaca aaacaaaaca aaaaccctga 2700
attagtggc tt当地aaaagg aaaaagcagg gctt当地aaaa agtagatcac acatcagtt 2760
agactcctgc tt当地tattta gtcaggtgt ct当地gattca gtctggagta ggcagagctt 2820
aagggtttt aagtccctgac ccaaagaaat gatctagcct gaaagtttag agcaaaggac 2880
taatgtttac tt当地aaagga atttcttgat tt当地aaaa aacttcattta aagtttaat 2940
ccccaaatgga caaattcata atctt当地taa tc当地tattac taaactttt aaaaatgtc 3000

ccaatttaca attaaataaa ttactttctc agtatattct ggtctggtca tggattgtgc 3060

atttccccc aaagatattc aaaattgtca attagagaat tttagggttt cagactcaga 3120

aaagtccctca cgccc 3135

<210> 28

<211> 3558

<212> DNA

<213> Homo sapiens

<400> 28

ctcagactga atacatggcc cactgtcgct ccagccatct caaatggaac gacctgttct 60

ctgaagtata tcttacagtg ct当地ctctcg aatccccctt gggaaatcta aaggctgaat 120

ccagccagct ttccatgct gc当地ggctcg gaaatcactg caagggttt tccagagaa 180

ccaaagtaag ataaatgaaa gatgctacac aattctgctg agggctctgt ctactcccc 240

tctcctgaaa cagctgttta ttcttcgac aggagttgaa accagcacct tccctttctc 300

tgagtcctgc ct当地ttctgc ggaagggagc tcaaaaagaac tttgttgtt tgcctttac 360

tctgggtga aagcggcagg aggtatgtga gatggtaaa tgatttgctt ctgccatgct 420

ggggtcacgg gtggatcgcc ct当地actctc ggtggccccc tcagtagttt tggaaagagga 480

ccaagtccct gt当地tccag cagtggacct ggaagaagga tgccggctca gggacttcac 540

tgagaaaata atgaatgtca aaggaaaagt aattctgtca atgctggttg tctcaactgt 600

gatcatgtg tttggaaat ttatcaaacag cacagaaggc tctttcttgt ggatatatca 660

ctcaaaaaac ccagaagttg atgacagcag tgctcagaag ggctgggtt ttctgagctg 720

57

gtttaacaat gggatccaca attatacaca agggaaagaa gacatagaca aaaaaaaaagg 780
aagagaggag accaaaggaa ggaaaatgac acaacagagc ttccggctatg ggactggtt 840
aatccaaaat ataatgatca ttacttggag gagttcataa catctgctaa taggtacttc 900
atggttggcc acaaagtcat atttacatc atggtgatg atgtctccaa gctgccgtt 960
atagagctgg gtcccttgca ttccttcaaa atgtttgagg tcaagccaga gaagaggtgg 1020
caagacatca gcatgatgcg tatgaagatc actggggagc acatcttggc ccacatccaa 1080
cacgaggtcg acttccttctt ctgcattggat gtggaccagg tcttccaaga ccattttggg 1140
gtggagaccc taggccagtc agtggctcag ctacaggctg gcggtaacaag gcagatccct 1200
atgactttac ctaggagagg tggaaagagt cagcaggata cattccattt ggccagggga 1260
tttttattac catgcagcca tttctggagg aacacccatt caggttctca acatcacccca 1320
ggagtgcctt aaggaaatcc tcctggacaa gaaaaatgac atagaagcca agtggcatga 1380
tgaaagccac ctaaacaagt atttccttct caataaaccc tctaaaatct tatccctaaa 1440
atactgctgg gattatcata taggcctgcc ttcaagatatt aaaactgtca agtgatcgtg 1500
gcagacaaaa gagtataatt tggtagaaa taatgtctga cttcaattt tgccagtaga 1560
tttctgaatt taagagagag aatattctgg ctacttcctc agaaaagtaa cacttaattt 1620
taacttc当地 aaatactaatacataat gaaacaccaa cagggcaaaa acataccatt cctccttgta 1680
acttggggct ttgtatgtg gaagaatgaa tctaggcataa tcagatataa attcccagtg 1740
atttcttatac tattctgggt ttggggaaa tactatcaac tgaaccaaaa ataacttgc 1800
ataggcagag ataaagccag aaacactcta cacatgccag atgacatctg gaaaaaggg 1860

58

tgctaaggga agcgttggc agcaagatata gattgttaagg ggttgtccct ttagttcaat 1920
gtctgcctat ttctgatggg tctaaagcaa catggagtta ctgtgcagca gaactctcag 1980
taaagacacc atttgccttg gcaatcctca aaaagcttca atagcagatt gcttcagacc 2040
atctgttagtc cgtcctttc tcatactggat gttgttggc ttctgtgcga aagattggtg 2100
gagtgtcccc gtagatatca tgggtgtgt tgatcagagt cccaaggaac ctgaatgagc 2160
caaggtgccc agcatgaagt caaaacaaag ccttgacatg agttgccat gaaatagcga 2220
agagagagtg gaagagagga gccaatcaact gtggggcagt gccaccctga gggcacttag 2280
ggtatgggt tggtgcttaa atacatcaca gatccaggtta ctgaatggga ggaagtgtgg 2340
gtgatttcca atctcattga ccctatgttc agggacttga acggaagatg tttcttgtgt 2400
tgccctaagtg gtattcagtc taccagactc tgcaacttgc atcttcaaata cttggtaaa 2460
gagatgtgga tgggtcaga gaaggcaaag gcctgcagtg gattgaagag gcttgcaagc 2520
agttctgttt ctaggatgtg ggcttcatca gaagacactc ggtcaccact tagctagtct 2580
aaacctcagg gttcctcagc ccatcataacc ccaacttggga ggactgacat caaggagtag 2640
actggagaaa cagccctccc atcaagtaac ctcttggtct ctcctgctcc atctgcacta 2700
tagaagtgtta ataatttagac atacttggca aaatggctaa ttgatttggt aacagaagca 2760
tgagccataa caatggaga tctagttatc atgactgaac agcttaacat tcaattccct 2820
tctctaaagag aagctgtgaa atcctacata ttattnaaag ttaacccaaat caatgtaaag 2880
ggagtttagga gacagtgtgt acctatgcac gtatatttat gtttgcttg tggccagtc 2940
tcggtcattt gttccattt tcaagcaatt tatttgaaga gccattgcac tagcctgatg 3000

59

tatactgcaa tgagcttctt tgataaaaatg aaacttaaat ttttctcgac catttcaccg 3060
 tgcctcctac ttcatttttt gccagaaaat ctcacatcca acaaaacaaa acaaaaaccc 3120
 tgaatttagtg ggcttgaaa aggaaaaagc agggcttga aaaagttagat cacacatcag 3180
 ttaagactcc tgcttctcta ttagtcaggt tgtctggat tcagtctgga gtaggcagag 3240
 cttaagggtt tttaagtccct gacccaaaga aatgatctag cctgaaagtt tagagcaaag 3300
 gactaatgtt tacttttaaa ggaatttctt gatTTTTTA aaaaacttca ttaaagttt 3360
 aatccccaat ggacaaattc ataatcttgt taatcgttat tactaaactt tttaaaaaat 3420
 gtcccaattt acaattaaat aaattacttt ctcagtatat tctggctgg tcatggattg 3480
 tgcatttcct cccaaagata ttcaaaattt tcaatttagag aatttttagt tttcagactc 3540
 agaaaagtcc tcacgccc 3558

<210> 29

<211> 852

<212> DNA

<213> Homo sapiens

<400> 29

gtggctgatc agagcgcgta gggcttcgcc gggggccggga gctgggcgcg gtcctgctca 60
 gcccagctca ccgcgcgccc gcccctcggcg ccctcggcgc cctggttctg cggatcagga 120
 gaaaataatg aatgtcaaag gaaaagtaat tctgtcaatg ctggttgtct caactgtgat 180
 cattgtgttt tggaaattt a tcaacagcac agaaggctct ttcttctggat tatatcactc 240
 aaaaaaccca gaagttgatg acagcagtgc tcagaagggc tggtggttgc tgagctgggt 300
 taacaatggg atccacaatt atcaacaagg ggaagaagac atagacaaag aaaaaggaag 360

agaggagacc aaaggaagga aaatgacaca acagagcttc ggctatggga ctggtttaat 420
ccaaacctga aggaatccga ataactaac tggactctgg ttttctgact cagtcctct 480
agaagacctg gactgagaga tcatgcgggtt aaggagtgtg taacaggcgg accacctgtt 540
gggactgcga gattctcaag gggaggact gggctcatt tctccatct cagcgcttag 600
caggatgacc tggatatagag caggaaactg gaaaatgtgg gtcagggat cagacactcc 660
agttgggtct tttatataaa ttaaatggca aaaggctcca tacccttctc cttcttcct 720
accctccact ttatctgcaa aatggaaatg atgataacac ccacttcata gaatggtcat 780
gaagatcaaa tgagagaata aaagtcaagc acttagcctc tggtgcacaa taagtattaa 840
ataagtataac ct 852

<210> 30
<211> 1232
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)..(118)
<223> This is exon 1

<220>
<221> misc_feature
<222> (119)..(207)
<223> This is exon 4

<220>

<221> .misc_feature
<222> (208)..(243)
<223> This is exon 5

<220>
<221> misc_feature
<222> (244)..(309)
<223> This is exon 6

<220>
<221> misc_feature
<222> (310)..(425)
<223> This is exon 7

<220>
<221> misc_feature
<222> (426)..(1232)
<223> This is exon 8b

<220>
<221> misc_feature
<222> (797)..(802)
<223> putative polyadenylation signal

<400> 30
gtggctgatc agagcgcgta gggcttcgcc gggccggga gctggcgcg gtcctgctca 60
gcccagctca ccgcgcgccg gccctcgccg ccctcgccgc cctggttctg cgatcaggaa 120
aaaaataatg aatgtcaaag gaaaagtaat tctgtcaatg ctggttgtct caactgtgat 180
cattgtgttt tggaaattta tcaacagcac agaaggctct ttcttgatca tatatcactc 240

62

aaaaaaccca gaagttgatg acagcagtgc tcagaagggc tggtggttcc tgagctggtt 300
taacaatggg atccacaatt atcaacaagg ggaagaagac atagacaaag aaaaaggaag 360
agaggagacc aaaggaagga aaatgacaca acagagcttc ggctatgggactggtt 420
ccaaacctga aggaatccga ataactaaac tggactctgg tttctgact cagtccttct 480
agaagacctg gactgagaga tcatgcgggtt aaggagtgtg taacaggcg accacctgtt 540
gggactgcga gattctcaag gggaaaggact gggcttcatt tctccatct cagcgcttag 600
caggatgacc tggtatagag cagggactg gggaaatgtgg gtcagggat cagacactcc 660
agttgggtct ttatataaa ttaaatggca aaaggctcca tacccttctc cttctttcct 720
accctccact ttatctgcaa aatggaaatg atgataacac ccacttcata gaatggtcat 780
gaagatcaa tgagagaata aaagtcaagc acttagcctc tggtgcacaa taagtattaa 840
ataagtatac ctattccctcc tttccctttt ttaaaaataa tattaccaaa tgtccagctt 900
atacacattnn acaagactta gctagtgccc tatgttagag ctactaaaag atctttgaca 960
agctaaaact aagatgcaat gaatgaggtg taacgaacaa gagagtttta agttcagaaa 1020
tggttacaga agtataagac agctgtgtgg gtgtttttt gttttgggtt tctgggttac 1080
aatctcgta ttcaacaaag atgggagttt tatagaacta aaagcaccat gtaagctact 1140
aaaaacaaca acaaaaaagg ctcatcattt ctcagtctga attgacaaaa atgccaatgc 1200
aaataaaaaat gattactttt tattttcaa cg 1232

<210> 31

<211> 1275

<212> DNA

<213> Homo sapiens

<400> 31

ctcagactga atacatggcc cactgtcgct ccagccatct caaatggaac gacctgttct	60
ctgaagtata tcttacagtg ctttctctcg aatccccttt gggaaatcta aaggctgaat	120
ccagccagct ttccatgct gcctggtctg gaaatcactg caagggttt tcccagagaa	180
ccaaagtaag ataaatgaaa gatgctacac aattctgctg agggctctgt ctactccccca	240
tctcctgaaa cagctgttta ttcttcgac aggagttgaa accagcacct tccctttctc	300
tgagtcctgc ctccctctgc ggaagggagc tcaaaagaac tttgttgtt tgccctttac	360
tctgggtga aagcggcagg aggtatgtga gatggtgaaa tgatttgctt ctgccatgct	420
ggggtcacgg gtggatcgcc ctaaactctc ggtggccccc tcagtagttt tggaagagga	480
ccaagtcctt gtctctccag cagtggacct ggaagaagga tgccggctca gggacttcac	540
tgagaaaaata atgaatgtca aaggaaaagt aattctgtca atgctggttt tctcaactgt	600
gatcattgtg tttggaaat ttatcaacag cacagaaggc tctttttgtt ggatatatca	660
ctcaaaaaac ccagaagttt atgacagcag tgctcagaag ggctgggtt ttctgagctg	720
gtttaacaat gggatccaca attatcaaca agggaaagaa gacatagaca aagaaaaagg	780
aagagaggag accaaaggaa ggaaaatgac acaacagagc ttccggctatg ggactggttt	840
aatccaaact tgaaggaatc cgaataacta aactggactc tggtttctg actcagtcct	900
tctagaagac ctggactgag agatcatgct gttaaggagt gtgtAACAGG cgaccaccc	960
gttggactg cgagattctc aagggaaagg actgggtctc atttctccca tctcagcgct	1020
tagcaggatg acctggatata gagcaggaa ctggaaatg tgggtcaggg gatcagacac	1080

tccagttggg tctttatat aaattaaatg gcaaaaggct ccataccctt ctccttctt 1140
cctaccctcc actttatctg caaaaatggga atgatgataa cacccacttc atagaatggt 1200
catgaagatc aaatgagaga ataaaagtca agcacttagc ctctggtgca caataagttat 1260
taaataagta tacct 1275

<210> 32
<211> 1655
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)..(272)
<223> This is exon 1a

<220>
<221> misc_feature
<222> (273)..(541)
<223> This is exon 2

<220>
<221> misc_feature
<222> (542)..(630)
<223> This is exon 4

<220>
<221> misc_feature
<222> (631)..(666)
<223> This is exon 5

<220>
<221> misc_feature
<222> (667)..(732)
<223> This is exon 6

<220>
<221> misc_feature
<222> (733)..(848)
<223> This is exon 7

<220>
<221> misc_feature
<222> (849)..(1655)
<223> This is exon 8h

<400> 32
ctcagactga atacatggcc cactgtcgct ccagccatct caaatggaac gacctgttct 60

ctgaagtata tcttacagtg ct当地ctctcg aatccccttt gggaaatcta aaggctgaat 120

ccagccagct ttccatgct gcctggcttg gaaatcactg caagggtttt tcccagagaa 180

ccaaagtaag ataaatgaaa gatgctacac aattctgctg agggctctgt ctactccccca 240

tctcctgaaa cagctgttta ttcttcgac aggagttgaa accagcacct tccctttctc 300

tgagtctgc ct当地tctgc ggaagggagc tcaaaaagaac tttgttgtt tgccctttac 360

tctgggtga aagcggcagg aggtatgtga gatggtgaaa tgatttgctt ctgccatgct 420

ggggtcacgg gtggatcgcc ct当地actctc ggtggccccc tcagtagttt tggaagagga 480

ccaagtccctt gtctctccag cagtggacct ggaagaagga tgccggctca gggacttcac 540

66

tgagaaaaata atgaatgtca aaggaaaagt aattctgtca atgctggttg tctcaactgt 600
gatcatttgtt ttttggaaat ttatcaacag cacagaaggc tctttcttgtt ggatatatca 660
ctcaaaaaac ccagaagttt atgacagcag tgctcagaag ggctgggtt ttctgagctg 720
gtttaacaat gggatccaca attatcaaca aggggaagaa gacatagaca aagaaaaagg 780
aagagaggag accaaaggaa ggaaaatgac acaacagagc ttccggctatg ggactggttt 840
aatccaaact tgaaggaatc cgaataacta aactggactc tggtttctg actcagtcct 900
tctagaagac ctggactgag agatcatgcg gtttaaggagt gtgttaacagg cgaccaccc 960
gttgggactg cgagattctc aaggggaagg actgggtctc atttctccca tctcagcgct 1020
tagcaggatg acctggtata gagcaggaa ctggaaatg tgggtcaggg gatcagacac 1080
tccagttggg tcttttatataaattaaatg gcaaaaggct ccataccctt ctccttctt 1140
cctaccctcc actttatctg caaaatggga atgatgataa cacccacttc atagaatgg 1200
catgaagatc aaatgagaga ataaaagtca agcacttagc ctctggtgca caataagtat 1260
taaataagta tacctattcc tcctttcct tttttaaaaa taatattacc aaatgtccag 1320
cttatacaca ttacaagac ttagcttagt ggctatgtta gagctactaa aagatcttt 1380
acaagctaaa actaagatgc aatgaatgag gtgttaacgaa caagagagtt ttaagttcag 1440
aaatggttac agaagtataa gacagctgtg tgggtgtttt ttgggttttgg tttctggtt 1500
tacaatctcg tcattcaaca aagatggag tttatagaa ctaaaagcac catgtaaatc 1560
actaaaaaaca acaacaaaaa aggctcatca tttctcagtc tgaattgaca aaaatgccaa 1620
tgcaaataaaa aatgattact ttttattttt caacg 1655

<210> 33
<211> 3322
<212> DNA
<213> Homo sapiens

<400> 33

gtggctgatc agagcgcgta gggcttcgcc gggccggga gctggcgcg gtcctgctca 60
gcccagctca ccgcgcgccc gccctcgccg ccctcgccgc cctggttctg cgatcaggaa 120
aaaaataatg aatgtcaaag gaaaagtaat tctgtcaatg ctggttgtct caactgtat 180
cattgtgttt tggaattta tcaacagcac agaaggctct ttcttgga tatatcactc 240
aaaaaaccca gaagttgatg acagcagtgc tcagaaggc tggtggttc tgagctgg 300
taacaatggg atccacaatt atcaacaagg ggaagaagac atagacaaag aaaaaggaag 360
agaggagacc aaaggaagga aaatgacaca acagagcttc ggctatggga ctggttaat 420
ccaaacctga aggaatccga ataactaaac tggactctgg ttttctgact cagtccttct 480
agaagacctg gactgagaga tcatgcgggtt aaggagtgtg taacaggcgg accacctgtt 540
gggactgcga gattctcaag gggaggact gggcttcatt tctccatct cagcgcttag 600
caggatgacc tgatataatg atcattactt ggaggagttc ataacatctg ctaataggta 660
cttcatgggtt gcccacaaag tcatattta catcatggtg gatgtgtct ccaagctgcc 720
gtttatagag ctgggtcctc tgcattcattt caaaatgttt gaggtcaagc cagagaagag 780
gtggcaagac atcagcatga tgcgtatgaa gatcaactggg gagcacatct tggcccacat 840
ccaacacgag gtcgacttcc tcttctgcat ggatgtggac caggtcttcc aagaccattt 900
tgggtggag accctaggcc agtcagtggc tcaagctacag gctggcgta caaggcagat 960

ccctatgact ttaccttagga gaggtggaaa gagtcagcag gatacattcc atttggccag 1020
gggattttta ttaccatgca gccatttctg gaggaacacc cattcagggtt ctcaacatca 1080
cccaggaggcgtt ctttaaggga atcctcctgg acaagaaaaa tgacatagaa gccaagtggc 1140
atgatgaaag ccacctaaac aagtattcc ttctcaataa accctctaaa atcttatccc 1200
taaaatactg ctgggattat catataggcc tgccttcaga tattaaaact gtcaagtgtat 1260
cgtggcagac aaaagagttt aatttggtaa gaaataatgt ctgacttcaa atttgccag 1320
tagatttctg aatttaagag agagaatatt ctggctactt cctcagaaaa gtaacactta 1380
attttaactt caaaaaatac taatgaaaca ccaacagggc aaaaacatac cattcctcct 1440
tgtaacttgg ggctttgtaa tgtggaaagaa tgaatctagg gcaatcagat ataaattccc 1500
agtgatttct tatctattct gggtttgggg gaaatactat caactgaacc aaaaataact 1560
tgtcataggc agagataaag ccagaaacac tctcacatg ccagatgaca tctggagaaa 1620
agggtgctaa gggaaacgtt tggcagcaag atatgattgt aaggggttgtt cccttgagtt 1680
caatgtctgc ctatttctga tgggtctaaa gcaacatgga gttactgtgc agcagaactc 1740
tcagtaaaga caccatttgc cttggcaatc ctcaaaaagc ttcaatagca gattgcttca 1800
gaccatctgt agtccgtcct tttctcatct ggatgttgg tggcttctgt gcgaaagatt 1860
ggtggagtgt cccagtagat atcatggtgg tgtgtgatca gagtcccaag gaacctgaat 1920
gagccaaagggt gcccaaggcatg aagtcaaaac aaagccttga catgagtttgc ccatgaaata 1980
gcgaagagag agtggaaagag aggagccat cactgtgggg cagtgccacc ctgagggcac 2040
ttagggtatg gggttgggtgc ttaaatacat cacagatcca ggtactgaat gggaggaagt 2100

gtgggtgatt tccaatctca ttgaccctat gttcagggac ttgaacggaa gatgtttctt 2160
gtgtgccta agtgttattc agtctaccag actctgcaac ttgcatttc aaatccttg 2220
taaagagatg tggatgggt cagagaaggc aaaggcctgc agtggattga agaggcttgc 2280
aagcagttct gttcttagga tgtggcttc atcagaagac actcggtcac cacttagcta 2340
gtctaaacct cagggttcct cagccatca taccccaact tggaggactg acatcaagga 2400
gttagactgga gaaacagccc tccccatcaag taacctcttg ttctctcctg ctccatctgc 2460
actatagaag tgtaataatt agacatactt ggcaaaatgg ctaattgatt tggtaacaga 2520
agcatgagcc ataacaatgg aagatctagt tatcatgact gaacagctt acattcaatt 2580
cccttctcta agagaagctg taaaatccta catatttattt aaagtttaacc aaatcaatgt 2640
aaaggagtt aggagacagt gtgtacctat gcacgtatat ttatgttttgc 2700
agtctcggtc atttgtttcc attttcaagc aatttatttg aagagccatt gcactagcct 2760
gatgtatact gcaatgagct tctttgataa aatgaaactt aaattttct cgaccatttc 2820
accgtgcctc ctacttcatt ttttgcaga aatctcaca tccaacaaaa caaaacaaaa 2880
accctgaatt agtgggcttt gaaaaggaaa aagcagggct ttgaaaaagt agatcacaca 2940
tcagtttaaga ctccctgcttc tctatttagtc aggttgtctt ggattcagtc tggagtaggc 3000
agagcttaag gtttttaag tcctgaccca aagaaatgat ctagcctgaa agtttagagc 3060
aaaggactaa tgtttacttt taaaggaatt tcttgatttt tttaaaaaac ttcattaaag 3120
tttaaatccc caatggacaa attcataatc ttgttaatcg ttattactaa actttttaaa 3180
aaatgtccca atttacaatt aaataaatta ctttctcagt atattctggt ctggcatgg 3240

attgtgcatt tcctccaaa gatattcaaa attgtcaatt agagaatttt aggtttcag 3300
actcagaaaa gtcctcacgc cc 3322

<210> 34
<211> 3745
<212> DNA
<213> Homo sapiens

<400> 34
ctcagactga atacatggcc cactgtcgct ccagccatct caaatggaac gacctgttct 60
ctgaagtata tcttacagtg ctttctctcg aatccccttt gggaaatcta aaggctgaat 120
ccagccagct ttccatgct gcctggctcg gaaatcactg caagggtttt tcccagagaa 180
ccaaagtaag ataaatgaaa gatgctacac aattctgctg agggctctgt ctactcccc 240
tctcctgaaa cagctgttta ttcttcgac aggagttgaa accagcacct tcccttctc 300
tgagtccctgc ctccctctgc ggaaggggagc tcaaaagaac tttgtgttt tgcctttac 360
tctgggtga aagcggcagg aggtatgtga gatggtgaaa tgatttgctt ctgccatgct 420
gggtcacgg gtggatcgcc ctaaactctc ggtggccccc tcaatgtttt tggaaagagga 480
ccaagtccctt gtctctccag cagtggacct ggaagaagga tgccggctca gggacttcac 540
tgagaaaata atgaatgtca aaggaaaagt aattctgtca atgctggttt tctcaactgt 600
gatcattgtg tttggaaat ttatcaacag cacagaaggc tctttttgt ggatatatca 660
ctcaaaaaac ccagaagttt atgacagcag tgctcagaag ggctgggtt ttctgagctg 720
gtttaacaat gggatccaca attatcaaca agggaaagaa gacatagaca aagaaaaagg 780

aagagaggag accaaaggaa ggaaaatgac acaaacagagc ttccggctatg ggactggttt 840
aatccaaact tgaaggaatc cgaataacta aactggactc tggtttctg actcagtcct 900
tctagaagac ctggactgag agatcatgcg gttaaggagt gtgtAACAGG cgaccacac 960
gttggactg cgagattctc aagggaaagg actgggtctc atttctccc tctcagcgct 1020
tagcaggatg acctgatata atgatcatta cttggaggag ttcataacat ctgctaata 1080
gtacttcatg gttggccaca aagtcatatt ttacatcatg gtggatgatg tctccaagct 1140
gccgtttata gagctggtc ctctgcattc cttcaaaatg tttgaggtca agccagagaa 1200
gaggtggcaa gacatcagca tgatgcgtat gaagatcaact ggggagcaca tcttggccca 1260
catccaaacac gaggtcgact tcctcttctg catggatgtg gaccaggct tccaaagacca 1320
ttttgggtg gagaccctag gccagtcagt ggctcagcta caggctggcg gtacaaggca 1380
gatccctatg actttaccta ggagaggtgg aaagagtcag caggatacat tccatttggc 1440
caggggattt ttattaccat gcagccattt ctggaggaac acccattcag gttctcaaca 1500
tcacccagga gtgcttaag ggaatcctcc tggacaagaa aaatgacata gaagccaagt 1560
ggcatgatga aagccaccta aacaagtatt tccttctcaa taaaccctct aaaatcttat 1620
ccctaaaaata ctgctggat tatcatata gctgccttc agatattaaa actgtcaagt 1680
gatcggtggca gacaaaagag tataatttgg ttagaaataa tgtctgactt caaattgtgc 1740
cagtagattt ctgaatttaa gagagagaat attctggcta cttcctcaga aaagtaacac 1800
ttaattttaa cttcaaaaaa tactaatgaa acaccaacag ggcaaaaaca taccattcct 1860
ccttgtaact tggggctttg taatgtggaa gaatgaatct agggcaatca gatataaatt 1920

72

cccaagtgatt tcttatctat tctgggtttg gggaaatac tatcaactga accaaaaata 1980
acttgtcata ggcagagata aagccagaaa cactctacac atgccagatg acatctggag 2040
aaaagggtgc taagggaaagc gtttggcagc aagatatgtat tgtaagggt tgcccttga 2100
gttcaatgtc tgccttatttc tcatgggtct aaagcaacat ggagttactg tgccagcaga 2160
ctctcagtaa agacaccatt tgccttggca atccctaaaaa agcttcaata gcagattgct 2220
tcagaccatc ttagtccgt cctttctca tctggatgtt gtttggcttc tgtgcgaaag 2280
attggtggag tgtcccagta gatatcatgg tgggtgtgtga tcagagtccc aaggaacctg 2340
aatgagccaa ggtgcccagc atgaagtcaa aacaaagcct tgacatgagt ttgccatgaa 2400
atagcgaaga gagagtggaa gagaggagcc aatcaactgtg gggcagtgcc accctgaggg 2460
cacttagggt atgggttgg tgcttaata catcacagat ccaggtactg aatgggagga 2520
agtgtgggt atttccaatc tcattgaccc tatgttcagg gacttgaacg gaagatgtt 2580
cttgtgttgc ctaagtggta ttcaacttac cagactctgc aacttgcac ttc当地atcct 2640
tggtaaagag atgtggatgg tgtcagagaa ggcaaaaggcc tgcagtgat tgaagaggct 2700
tgcaagcagt tctgtttcta ggatgtggc ttcatcagaa gacactcggt caccacttag 2760
ctagtctaaa cctcagggtt cctcagccca tcataccccca acttggagga ctgacatcaa 2820
ggagtagact ggagaaacag ccctccatc aagtaacctc ttgttcttc ctgctccatc 2880
tgcactatag aagtgtata attagacata ctggcaaaa tggctaattt atttggtaac 2940
agaagcatga gccataacaa tggaaagatct agttatcatg actgaacagc ttaacattca 3000
atcccttct ctaagagaag ctgtgaaatc ctacatatta tttaaagtta accaaatcaa 3060

73

tgtaaagggga gtaggagac agtgttacc tatgcacgta tatttatgtt ttgcttgtgt 3120
tccagtctcg gtcatttgtt tccattttca agcaatttat ttgaagagcc attgcactag 3180
cctgatgtat actgcaatga gcttcttga taaaatgaaa cttaaatttt tctcgaccat 3240
ttcacccgtgc ctcctacttc atttttgcc agaaaatctc acatccaaca aaacaaaaaca 3300
aaaaccctga attagtggc tttgaaaagg aaaaagcagg gctttgaaaa agtagatcac 3360
acatcagtttta agactcctgc ttctctatta gtcaggttgc cttggattca gtctggagta 3420
ggcagagctt aagggtttt aagtccgtac ccaaagaaat gatctagcct gaaagtttag 3480
agcaaaggac taatgtttac ttttaaagga atttcttgat tttttaaaa aacttcatttta 3540
aagtttaaat ccccaatgga caaattcata atcttggtaa tcgttattac taaactttt 3600
aaaaaatgtc ccaatttaca attaaataaa ttactttctc agtatattct ggtctggtca 3660
tggattgtgc atttcctccc aaagatattc aaaattgtca attagagaat tttaggttt 3720
cagactcaga aaagtccctca cgccc 3745

<210> 35
<211> 244
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)..(60)
<223> 5' flanking sequence

<220>
<221> misc_feature

<222> (61)..(178)

<223> human untranslated exon 1

<220>

<221> Intron

<222> (179)..(244)

<400> 35

agcccggccg gccggcccac gggcgggagg acgcgcctcc gctcgggcgg aggcggcg 60

gtggctgatc agagcgcgta gggcttcgcc gggccggga gctgggcg 120

gcccagctca ccgcgcgccc gccctcggcg ccctcggcgc cctggttctg cgatcagg 180

gggtcccgcg gggagccgcc caggtccccg gaggccacga gcaggacacg gacggggggc 240

nnnn 244

<210> 36

<211> 217

<212> DNA

<213> Homo sapiens

<220>

<221> Intron

<222> (1)..(60)

<220>

<221> misc_feature

<222> (61)..(149)

<223> human untranslated exon 4

<220>

<221> Intron

<222> (150)..(217)

<400> 36
ctcttgaagt tcattgattt aatctgttct cttttttct cccctttct ttttcctag 60

gagaaaataaa tgaatgtcaa aggaaaagta attctgtcaa tgctgggtgt ctcaactgtg 120

atcattgtgt ttggaaatt tatcaacagg taattatgaa acatgatgaa gtgatgtgga 180

tgaaaataact gctttgattc tatcctacta gatatnnn 217

<210> 37
<211> 165
<212> DNA
<213> Homo sapiens

<220>
<221> Intron
<222> (1)..(60)

<220>
<221> misc_feature
<222> (61)..(96)
<223> human untranslated exon 5

<220>
<221> Intron
<222> (97)..(165)

<400> 37
aatcgccctt ctcagaatta aaagtaacat gatatgtttt tatttctttt ttgcttttag 60

cacagaaggc tctttcttgtt ggatatatca ctcaaagtgc tttgaattct agatttctag 120

gggatgtttc ccacagccac tctggcaccc cctacagtcc annnn 165

<210> 38
<211> 193
<212> DNA
<213> Homo sapiens

<220>
<221> Intron
<222> (1)..(60)

<220>
<221> misc_feature
<222> (61)..(126)
<223> human untranslated exon 6

<220>
<221> Intron
<222> (127)..(193)

<400> 38
accctaagtt tggggacacc acattttcta aaaatatttg taaactttt catttcttag 60

aaacccagaa gttgatgaca gcagtgctca gaagggtctgg tggtttctga gctggtttaa 120

caatggtaa ggcggatcag acagcagtcg gtgttgccc acccgctgg tgcttgaga 180

gggtccnnnn nnn 193

<210> 39
<211> 242
<212> DNA
<213> Homo sapiens

<220>
<221> Intron
<222> (1)..(60)

<220>
<221> misc_feature
<222> (61)..(176)
<223> human untranslated exon 7

<220>
<221> Intron
<222> (177)..(242)

<400> 39

tctttgacca ccgcaatcac cttccctgcc ttacctggtt tactttccct ttgtacttag 60

gatccacaat tatcaacaag gggagaaga catagacaaa gaaaaaggaa gagaggagac 120

caaaggaagg aaaatgacac aacagagctt cgctatggg actggttaa tccaaagtaa 180

gaaaagcggc gtcactccct gtgcagcaaa tccatggccc tgcagggggt ggtgtggcnn 240

nn 242

<210> 40
<211> 487
<212> DNA
<213> Homo sapiens

<220>
<221> Intron
<222> (1)..(60)

<220>
<221> misc_feature
<222> (61)..(487)
<223> a version of human untranslated exon 8h

<400> 40

atagaatatt ttaattttta attcaacata aatttttaag ggtgctgttt tttcttccag 60
cttgaaggaa tccgaataac taaaactggac tctggtttc tgactcagtc cttctagaag 120
acctggactg agagatcatg cggttaagga gtgtgtaaca ggccggaccac ctgttgggac 180
tgcgagattc tcaagggaa ggactgggtc tcatttctcc catctcagcg cttagcagga 240
tgacctggta tagagcaggg aactggaaaa tgtgggtcag gggatcagac actccagttg 300
ggtcttttat ataaattaaa tggcaaaagg ctccataccc ttctccttct ttcttaccct 360
ccactttatc tgcaaaaatgg gaatgatgat aacacccact tcatagaatg gtcatgaaga 420
tcaaatgaga gaataaaaagt caagcactta gcctctggtg cacaataagt attaaataag 480
tataacct 487

<210> 41
<211> 454
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)..(380)
<223> a version of the human untranslated exon 8h

<220>
<221> Intron
<222> (381)..(454)

<400> 41
attcctcctt ttcccttttt aaaaataata ttaccaaatg tccagcttat acacattac 60
aagacttagc tagtgggcta tgttagagct actaaaagat ctttgacaag ctaaaaactaa 120

gatgcaatga atgaggtgta acgaacaaga gagtttaag ttcagaaatg gttacagaag 180
tataagacag ctgtgtgggt gtttttggt ttttggttc tggttacaa tctcgtcatt 240
caacaaagat gggagttta tagaactaaa agcaccatgt aagctactaa aaacaacaac 300
aaaaaggct catcatttct cagtctgaat tgacaaaaat gccaatgcaa ataaaaatga 360
ttactttta ttttcaacg ttgttgttt atttatttat ttcgagatgg agtttcactc 420
ttgttgccct ggctggagtg cagtggcgcn nnnn 454

<210> 42
<211> 2848
<212> DNA
<213> Homo sapiens

<220>
<221> Intron
<222> (1)..(65)

<220>
<221> misc_feature
<222> (66)..(2676)
<223> human untranslated exon 9

<220>
<221> misc_feature
<222> (2677)..(2848)
<223> an inter-gene sequence

<400> 42
ttcagcttgt ggtttcttc aggaatccca gaggataaat gtttgcttt tcttctttgt 60

80

ttcagatata atgatcatta cttggaggag ttcataacat ctgctaata gtaacttcatg 120
gttggccaca aagtcatatt ttacatcatg gtggatgtatg tctccaagct gccgtttata 180
gagctgggtc ctctgcattc cttcaaaatg tttgaggtca agccagagaa gaggtggcaa 240
gacatcagca ttagatcgat gaagatcaact gggagcaca tcttggccca catccaacac 300
gaggtcgact tcctcttctg catggatgtg gaccaggct tccaagacca ttttggggtg 360
gagaccttag gccagtcagt ggctcagcta caggctggcg gtacaaggca gatccctatg 420
actttaccta ggagaggtgg aaagagtcag caggatacat tccatttggc caggggattt 480
ttattaccat gcagccattt ctggaggaac acccattcag gttctcaaca tcacccagga 540
gtgcttaag ggaatcctcc tggacaagaa aaatgacata gaagccaagt ggcatgtga 600
aagccaccta aacaagtatt tccttctcaa taaaccctct aaaatcttat ccctaaaata 660
ctgctggat tatcatata gctgccttc agatattaa actgtcaagt gatcgtggca 720
gacaaaagag tataatttgg ttagaaataa tgtctgactt caaatgtgc cagtagattt 780
ctgaatttaa gagagagaat attctggcta cttcctcaga aaagtaaacac ttaattttaa 840
cttcaaaaaa tactaatgaa acaccaacag ggcaaaaaca taccattcct ccttgtaact 900
tggggctttaatgtggaa gaatgaatct agggcaatca gatataaattt cccagtgttt 960
tcttatctat tctgggtttg gggaaatac tatcaactga accaaaaata acttgtcata 1020
ggcagagata aagccagaaa cactctacac atgccagatg acatctggag aaaagggtgc 1080
taagggaaagc gttggcagc aagatatgtatg tggtaagggt tgcccttga gttcaatgtc 1140
tgcctatttc ttaggggtct aaagcaacat ggagttactg tgcagcagaa ctctcagtaa 1200

agacaccatt tgccttggca atcctcaaaa agcttcaata gcagattgct tcagaccatc 1260
tgtagtcgt cctttctca tctggatgtt gttggcttc tgtgcgaaag attggtgag 1320
tgtcccagta gatatcatgg tggtgtgtga tcagagtccc aaggaacctg aatgagccaa 1380
ggtgcccagc atgaagtcaa aacaaaggct tgacatgagt ttgccatgaa atagcgaaga 1440
gagagtggaa gagaggagcc aatcactgtg gggcagtgcc accctgaggg cacttaggg 1500
atggggttgg tgcttaata catcacagat ccaggtactg aatgggagga agtgtgggtg 1560
atttccaatc tcattgaccc tatgttcagg gacttgaacg gaagatgttt ctgtgttgc 1620
ctaagtggta ttcatgtctac cagactctgc aacttgcattt ttcaaattcct tggtaaagag 1680
atgtggatgg tgtcagagaa ggcaaaggcc tgcagtggat tgaaggaggct tgcaaggcgt 1740
tctgtttcta ggatgtgggc ttcatcagaa gacactcggt caccacttag ctgtctaaa 1800
cctcagggtt cctcagccca tcataccca acttggagga ctgacatcaa ggagtagact 1860
ggagaaaacag ccctcccatc aagtaacctc ttgttctctc ctgctccatc tgcactata 1920
aagtgtata attagacata cttggaaaaa tggctaattt atttggtaac agaagcatga 1980
gccataacaa tggaaagatct agttatcatg actgaacagc ttaacattca attcccttct 2040
ctaagagaag ctgtgaaatc ctacatatta tttaaagtta accaaatcaa tgtaaaggga 2100
gttaggagac agtgtgtacc tatgcacgta tatttatgtt ttgcttgtt tccagtctcg 2160
gtcatttggc tccatattca agcaattttt ttgaagagcc attgcactag cctgatgtat 2220
actgcaatga gcttcttga taaaatgaaa cttaaatttt tctcgaccat ttcaccgtgc 2280
ctcctacttc atttttgcc agaaaatctc acatccaaca aaacaaaaca aaaaccctga 2340

82

attagtggc tttgaaaagg aaaaagcagg gcttgaaaa agtagatcac acatcagtt 2400
agactcctgc ttctctatta gtcaggttgt ctggattca gtctggagta ggcagagctt 2460
aagggtttt aagtccgtac ccaaagaat gatctagcct gaaagtttag agcaaaggac 2520
taatgtttac ttttaaagga atttcttgat tttttaaaa aacttcatttta aagtttaaat 2580
ccccaatgga caaattcata atcttggtaa tcgttattac taaactttt aaaaaatgtc 2640
ccaatttaca attaaataaa ttactttctc agtatattct ggtctggtca tggattgtgc 2700
attcctccc aaagatattc aaaattgtca attagagaat ttttaggtttt cagactcaga 2760
aaagtcctca cgcccttctg aaaatgtgtc cactattaca gaaatagaac agacttggga 2820
ttcccaaatt tttgtttgtt tttnnnnn 2848

<210> 43
<211> 2303
<212> DNA
<213> Rhesus monkey

<220>
<221> misc_feature
<222> (1)..(44)
<223> This is exon 1

<220>
<221> misc_feature
<222> (45)..(159)
<223> This is exon 2

<220>
<221> misc_feature

<222> (160)..(278)
<223> This is exon 3

<220>
<221> misc_feature
<222> (279)..(367)
<223> This is exon 4

<220>
<221> misc_feature
<222> (368)..(403)
<223> This is exon 5

<220>
<221> misc_feature
<222> (404)..(469)
<223> This is exon 6

<220>
<221> misc_feature
<222> (470)..(584)
<223> This is exon 7

<220>
<221> misc_feature
<222> (585)..(2260)
<223> This is exon 9

<220>
<221> misc_feature
<222> (543)..(545)
<223> This is an early stop codon

<220>
<221> misc_feature
<222> (1276)..(1278)
<223> This is a stop codon

<220>
<221> misc_feature
<222> (1225)..(1260)
<223> This is a polyadenylation signal

<400> 43
gctcgctgcg cggccgtcct gggtgccagg gttctgcggta tcaggagttt aaaccagcat 60
cttcccttca tctgagtcct gcctccttct gcagaaggga gctcaaaaga actttgttgt 120
tttgcctttt actctggggt gaaagcaaca gacgataagg atctcactct gtcgccccaaag 180
ctggaggtgca gtggcttgat tacagctcac tgttagcctgg accttccaag gctctgggtg 240
atcttcctac ctcagcttcc ccagtagctg gactacagga gaaaataatg aatgtcaaag 300
gaaaagtaat tctgtcaatg ctgggtgtct caactgtgat cattgtgttt tggaatata 360
tcaatagccc agaaggttct ttcttgggta tgtatcgctc aaaaaaccca gaggttgatg 420
acagcagtgc tcagaagagc tggtggttc cgagctggtt taacaatggg atccacaatc 480
atcaacaaga ggaagaagac atagacaaaa aagaggaaga gaggagacca aagaaaggaa 540
gatgacacaa cagagcttcg gctatggac tgatttaatc caaaatatat tgagcattac 600
ttggaagagt tcataacacc tgctaatagg tacttcaagg tcggccacaa agtcatattt 660
tacattatacg tggatgatgt ctccaaagggtg ctgtttatag agctgggtcc tctgcattcc 720

ttaaaagtgt ttgaggtcaa gccagagaag aggtggcaac acatcagcat gatgcctgtg 780
aagatcatca gggagcacat cttggccac atccaacacg aggtcgactt cctcttctgc 840
atggatgttag accaggctt ccaagacaat ttggggtaa agaccctagg tcagtcagtg 900
gctcagctac agccctggtg gtacaaggca gatcctgatg actttaccta ggagaggcag 960
aaagagtcag cagcatgcat tccatttggc caggaggatt tttattacca cacagccatt 1020
tttggagggaa caccattca ggttctcaac atccccagg agtgcttaa gagaatcctc 1080
ctggaaaaga aaaatgacat agaagctgag tggcatgatg aaagccacct aaaccagtat 1140
ttccttctca acaaaccctc taaaatctta tccctagaat actgctggta 1200
agcctgcctt cagatattaa aactgtcaag cggtcggtgc agacaaaaga gtataatttg 1260
gttagaaata tcatactgact tcaaatttgtc ccagtagatt tctgaatttg agagaggagt 1320
attctggctg ctccctcaga aaagtaaacac ttaattttaa gttaaaaaaaaa atactaatga 1380
aacaccaaca tggcaaacac ataccattcc ttcttgtaac ttgaggctt gtaatgtggg 1440
agaatgaatc taggtaatc agatgtaaat tcccagtgtat ttcttatcta ttttgggttt 1500
gggggaaata ctatcaactg aaccaaaaag aacttgtcat aggcaaagat aaagccagaa 1560
acactctaca catgccacat aacatctggaa gaaaagggtg ctaaggaaag cgtttggcag 1620
caagatataa ttgttaagggg ttgtcccttg agttcaatgc ctgccttattt ccaatggatc 1680
taaaacaacg tgaagttact gtgcagcaga gctctcagta aggacaccat ttgccttggc 1740
aatcctcaaa attcttcaat agcagattgt ttcaggccat ctgttagtctg tcctttctc 1800
atcaggatgt ttttggctt ctgtgcgaaa aattgggtgaa gtgtcctgggt agatattgaa 1860

actaggcctc atatagaaaa aattaacacc aggtggctct ggatagagtc ccgcctgcc 1920
tcgatgagga cccaccctga tagggtcccc ccctgccaat tccgagaaac aacctcatgg 1980
ggtcccaccc tgccaattcc gggggtcccc ccctgcctcg aagttcccg aatcaacaac 2040
tccagggaaaa aacctcataa ggtcctgctc taaccaatta gcataagacg ccttgctcag 2100
gccatagcta gacccaatca ttttgcgcct taagcttgt ttgaatttcg cgccctaagc 2160
tgtgtttgaa cttgtgtttg cctatataaa cagcctgtaa caagcagtcg gggtcccagg 2220
gccaacttag agcttgggac cctagcgcgc tagtaataaa taactctctg ctgcgaaaaaa 2280
aaaaaaaaaaa aaaaaaaaaaa aaa 2303

<210> 44
<211> 2630
<212> DNA
<213> Rhesus monkey

<220>
<221> misc_feature
<222> (1)..(44)
<223> This is exon 1

<220>
<221> misc_feature
<222> (45)..(159)
<223> This is exon 2

<220>
<221> misc_feature
<222> (160)..(278)

<223> This is exon 3

<220>

<221> misc_feature

<222> (279)..(367)

<223> This is exon 4

<220>

<221> misc_feature

<222> (368)..(403)

<223> This is exon 5

<220>

<221> misc_feature

<222> (404)..(469)

<223> This is exon 6

<220>

<221> misc_feature

<222> (470)..(584)

<223> This is exon 7

<220>

<221> misc_feature

<222> (585)..(911)

<223> This is exon 8

<220>

<221> misc_feature

<222> (912)..(2587)

<223> This is exon 9

<220>
<221> misc_feature
<222> (288)..(290)
<223> This is a putative start codon

<220>
<221> misc_feature
<222> (543)..(545)
<223> This is a putative stop codon

<220>
<221> misc_feature
<222> (1603)..(1605)
<223> This is a putative stop codon

<220>
<221> misc_feature
<222> (2582)..(2587)
<223> polyadenylation signal

<400> 44
gctcgctgcg cgccggtcct gggtgccagg gttctgcgga tcaggagttg aaaccagcat 60

cttcccttca tctgagtcct gcctccttct gcagaaggga gctcaaaaga actttgttgt 120

tttgcctttt actctggggta gaaagcaaca gacgataagg atctcactct gtcgccaaag 180

ctggagtgca gtggcttgat tacagctcac tgttagcctgg accttccaag gctctgggtg 240

atcttcctac ctcagcttcc ccagtagctg gactacagga gaaaataatg aatgtcaaag 300

gaaaagtaat tctgtcaatg ctgggttgtct caactgtgat cattgtgttt tggaatata 360

89

tcaatagccc agaagggtct ttcttgggga tgtatcgctc aaaaaaccca gaggttgatg 420
acagcagtgc tcagaagagc tggtggttc cgagctggtt taacaatggg atccacaatc 480
atcaacaaga ggaagaagac atagacaaa aagaggaaga gaggagacca aagaaaggaa 540
gatgacacaa cagagcttcg gctatggac tgatttaatc caaagaaacg cccagaggtg 600
gtgagagtga ccagatggaa ggcaccgggtt gtgtggaaag gcacttacaa caagccatc 660
ctagggaaatt attatgc当地 acagaaaatt acggtggat tgaaggctt tgctattgga 720
agtgggtgtc actgatgaaa ctgtccttga ctatttcttgc ttccactgtc aagacattt 780
tgtggagact cctgaactga tggaggccag ccatgattt ttgatttatt agatagaaga 840
atgtttcat ggaactgttt tagtctcctt tctgctgagg ccctaaaatg ctgagaacaa 900
aataagagta gatatattga gcattacttg gaagagttca taacacctgc taataggtac 960
ttcaagggtcg gccacaaaatg catattttac attatagtgg atgatgtctc caaggtgctg 1020
tttatacgtgc tgggtcctct gcattcctta aaagtgtttg aggtcaagcc agagaagagg 1080
tggcaacaca tcagcatgtat gcctgtgaag atcatcaggg agcacatctt ggccccatc 1140
caacacgagg tcgacttcctt cttctgtcatg gatgttagacc aggtcttcca agacaattt 1200
gggggtgaaga cccttaggtca gtcagtggtc cagctacagc cctgggtgta caaggcagat 1260
cctgtatgtact ttaccttagga gaggcagaaa gagtcagcag catgcattcc atttggccag 1320
gaggatttt attaccacac agccatttt ggaggaacac ccattcaggt tctcaacatc 1380
ccccaggagt gctttaagag aatcctcctg gaaaagaaaa atgacataga agctgagtgg 1440
catgtgaaa gccacctaataa ccagtatttc cttctcaaca aaccctctaa aatcttatcc 1500

90

ctagaatact gctgggatta tcataatcagc ctgccttcag atattaaaac tgtcaagcgg 1560
tcgtggcaga caaaagagta taatttggtt agaaatatca tctgacttca aattgtgcc 1620
gtagatttct gaatttgaga gaggagtatt ctggctgctt cctcagaaaa gtaacactta 1680
attttaagtt aaaaaaaaata ctaatgaaac accaacatgg caaacacata ccattccttc 1740
ttgttaacttg aggcttgta atgtggaga atgaatctag ggtaatcaga tgtaaattcc 1800
cagtgatttc ttatctatTT tgggtttggg gggaaatacta tcaactgaac caaaaagaac 1860
ttgtcatagg caaagataaa gccagaaaca ctctacacat gccacataac atctggagaa 1920
aagggtgcta aggaaagcgt ttggcagcaa gatatgatttga taaggggttg tcccttgagt 1980
tcaatgcctg cctatttcca atggatctaa aacaacgtga agttactgtg cagcagagct 2040
ctcagtaagg acaccatttg ccttggcaat cctcaaaatt cttcaatagc agattgttcc 2100
aggccatctg tagtctgtcc ttttctcatc agatgttgtt tggtttctg tgcgaaaaat 2160
tggtggagtg tcctggtaga tattgaaact aggccctata tagaaaaaat taacaccagg 2220
tggctctgga tagagtcccg ccctgcctcg atgaggaccc accctgatag ggtcccaccc 2280
tgccaattcc gagaaacaac ctcattgggt cccaccctgc caattccggg ggtcccaccc 2340
tgcctcgaag ttcccgaaat caacaactcc agaaaaaaac ctcataaggt cctgctctaa 2400
ccaatttagca taagacgcct tgctcaggcc atagctagac ccaatcattt tgccgcctaa 2460
gcttttttgc aatttcgcgc cctaagctgt gttgaactt gtgtttgcct atataaacag 2520
cctgtaacaa gcagtcgggg tcccagggcc aacttagagc ttgggaccct agcgcgctag 2580
taataaataaa ctctctgctg cgaaaaaaaaaaaaaaaaaaaaaaaa 2630

<210> 45
<211> 35
<212> DNA
<213> Artificial/Unknown

<220>
<221> misc_feature
<222> ()...()
<223> Antisense primer for cloning porcine exon 4

<400> 45
ctgttgatgt attcccaaaa cacaaccatt acagt

35

<210> 46
<211> 27
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()...()
<223> Antisense primer for cloning porcine exon 4

<400> 46
agacaaggcag cattgacaga accactc

27

<210> 47
<211> 25
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature

<222> ()..()

<223> Antisense primer for cloning porcine exon 2

<400> 47

ctcatcctct gcttctctcc cccca

25

<210> 48

<211> 26

<212> DNA

<213> artificial sequence

<400> 48

ccccccagag taaaaggcga aacaag

26

<210> 49

<211> 25

<212> DNA

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

<223> Sense primer for cloning porcine exon 2

<400> 49

aacgcagcac ctcccttcc tccca

25

<210> 50

<211> 25

<212> DNA

<213> artificial sequence

<400> 50

cttgtttcgc cttttactct ggggg

25

<210> 51

<211> 23

<212> DNA

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

<223> Sense primer for cloning porcine exon 1

<400> 51

gccactgttc cctcagccga gga

23

<210> 52

<211> 24

<212> DNA

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

<223> Sense primer for cloning porcine exon 1

<400> 52

cgagcgcacc cagcttctgc cgat

24

<210> 53

<211> 24

<212> DNA

<213> artificial sequence

94

<220>
<221> misc_feature
<222> ()..()
<223> Antisense primer for cloning porcine exon 1

<400> 53
tgcgctcgaa gatggccctc tcct

24

<210> 54
<211> 24
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> Antisense primer for cloning porcine exon 1

<400> 54
ggcgtcctcg gctgagggaa cagt

24

<210> 55
<211> 28
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> Sense primer for cloning porcine exon 1A

<400> 55
cagaacaact tctgaaggct aaaggatg

28

<210> 56
<211> 27
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> Sense primer for cloning porcine exon 1

<400> 56
caaatggtgg atcggacctc ccaggct

27

<210> 57
<211> 27
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> Sense primer for cloning porcine exon 1

<400> 57
agtactgggt gatagacccc actccac

27

<210> 58
<211> 25
<212> DNA
<213> artificial sequence

<220>

96

<221> misc_feature
<222> ()..()
<223> Sense primer for cloning porcine exon 1

<400> 58
gcgcaggct ccggggcccc tccct

25

<210> 59
<211> 27
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> Sense primer for cloning porcine exon 9

<400> 59
ctgggattat catataggca tgtctgt

27

<210> 60
<211> 27
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> Sense primer for cloning porcine exon 9

<400> 60
agagtattac tctggctact tctccag

27

<210> 61
<211> 27
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()...()
<223> primer for identifying 5' flanking region of murine exon 1

<400> 61
ctgagagcgc gaggtcttca gcagaat 27

<210> 62
<211> 28
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()...()
<223> primer for identifying 5' flanking region of murine exon 1

<400> 62
cttctcattc caagaagagt cttacaag 28

<210> 63
<211> 27
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature

<222> ()..()

<223> primer for identifying 3' flanking region of murine exon 1

<400> 63

cctgcctttt cttagctggc tgacacc

27

<210> 64

<211> 27

<212> DNA

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

<223> primer for identifying 3' flanking region of murine exon 1

<400> 64

cttgttagact cttcttgaa tgagaag

27

<210> 65

<211> 27

<212> DNA

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

<223> primer for identifying 5' flanking region of murine exon 2

<400> 65

catcgtcagc tgtgttccct ccaaagc

27

99

<210> 66

<211> 27

<212> DNA

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

<223> primer for identifying 5' flanking region of murine exon 2

<400> 66

aaagcaaccg agcttctgtc gagctct

27

<210> 67

<211> 38

<212> DNA

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

<223> primer for identifying murine exons 2 and 3

<400> 67

gtacacctcct ttccctctgct gagccctgcc tccttcgg

38

<210> 68

<211> 35

<212> DNA

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

100

<223> primer for identifying murine exons 2 and 3

<400> 68

agatcttgag gatccaagac ttgtttctga ctgg

35

<210> 69

<211> 34

<212> DNA

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

<223> primer for identifying murine exons 3 and 4

<400> 69

gctgactttg aactcaagag atctgcttta cccc

34

<210> 70

<211> 28

<212> DNA

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

<223> primer for identifying murine exons 3 and 4

<400> 70

ctgttgacat attcccaaaa cacgacaa

28

<210> 71

101

<211> 30
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> primer for identifying murine exons 4 and 5

<400> 71
gtcaaggaa aagtaatcct gttgatgctg 30

<210> 72
<211> 27
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> primer for identifying murine exons 4 and 5

<400> 72
tatccacaag aaagagccgt ctgggct 27

<210> 73
<211> 27
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> primer for identifying murine exons 5 and 6

<400> 73
agccccagacg gctctttctt gtggata 27

<210> 74
<211> 34
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> primer for identifying murine exons 5 and 6

<400> 74
ccagcttggg aaccaccagt ccttctgccca tctg 34

<210> 75
<211> 27
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> primer for identifying murine exons 6 and 7

<400> 75
ttccagaggt tggtgagaac agatggc 27

<210> 76
<211> 33

103

<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> primer for identifying murine exons 6 and 7

<400> 76
gcgatctcca tttctaccct tttctctccg tcc 33

<210> 77
<211> 28
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> primer for identifying murine exon 7

<400> 77
caagaagaca acgtagaagg acggagag 28

<210> 78
<211> 27
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> primer for identifying murine exon 7

<400> 78

tcgcattgaa gagcctcagc tatggga

27

<210> 79

<211> 27

<212> DNA

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

<223> primer for identifying exon 8

<400> 79

ccacagttag tttctgtgtg gcgatgt

27

<210> 80

<211> 28

<212> DNA

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

<223> primer for identifying murine exon 8

<400> 80

agagctgtgt cataagtgcc ttccccaca

28

<210> 81

<211> 27

<212> DNA

105

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

<223> primer for identifying murine exon 8

<400> 81

gatgttttga cagtgacccc gtggaag

27

<210> 82

<211> 28

<212> DNA

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

<223> primer for identifying murine exon 8

<400> 82

tgtgggaagg cacttatgac acagctct

28

<210> 83

<211> 27

<212> DNA

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

<223> primer for identifying murine exon 9

106

<400> 83

agagggttca ggtgcacgac aggcatc

27

<210> 84

<211> 27

<212> DNA

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

<223> primer for identifying murine exon 8

<400> 84

gtacatgtca gcagactcca gaaagtc

27

<210> 85

<211> 27

<212> DNA

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

<223> primer for identifying 3' flanking region of murine exon 9

<400> 85

gactttctgg agtctgctga catgtac

27

<210> 86

<211> 27

<212> DNA

<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> primer for identifying 3' flanking region of murine exon 9

<400> 86
gatgcctgtc gtgcacctga accctct

27

<210> 87
<211> 27
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> primer for identifying 3' flanking region of murine exon 9

<400> 87
aggccattgc accatcttgg tgaacag

27

<210> 88
<211> 28
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> primer for identifying 3' flanking region of murine exon 9

<400> 88

108

gatcttacct ttgtccacag ggctctac

28

<210> 89

<211> 27

<212> DNA

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

<223> primer for obtaining murine promoter

<400> 89

ccaatgcattttcccaagt gggctct

27

<210> 90

<211> 27

<212> DNA

<213> artificial sequence

<220>

<221> misc_feature

<222> ()..()

<223> primer for isolation of transcription initiation site

<400> 90

cccagaacag atctgactgc ctcttgc

27

<210> 91

<211> 27

<212> DNA

<213> artificial sequence

109

<220>
<221> misc_feature
<222> ()...()
<223> primer for isolation of transcription initiation site

<400> 91
agttttgctt gtctgggccca ctatcgg

27

<210> 92
<211> 27
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()...()
<223> primer for isolation of transcription initiation site

<400> 92
gactggagag agtgctgtcc tccttgc

27

<210> 93
<211> 29
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()...()
<223> primer for cloning Rhesus alpha 1,3 GT

<400> 93
gaggtaaagc cagagaagag gtggcaaca

29

<210> 94
<211> 30
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> primer for cloning Rhesus alpha 1,3 GT

<400> 94
gacttcctct tctgcattgg a tgttagaccag

30

<210> 95
<211> 29
<212> DNA
<213> artificial sequence

<220>
<221> misc_feature
<222> ()..()
<223> primer for cloning Rhesus alpha 1,3 GT

<400> 95
atgtcgagaa cctgaatggg tgttcctcc

29

<210> 96
<211> 30
<212> DNA
<213> artificial sequence

<220>

111

<221> misc_feature
<222> ()..()
<223> primer for cloning Rhesus alpha 1,3 GT

<400> 96

ctggccaaat ggaatgcattg ctgctgactc

30

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
3 May 2001 (03.05.2001)

PCT

(10) International Publication Number
WO 01/30992 A3(51) International Patent Classification⁷: C12N 15/54.

5/10, 15/85, 9/10. A01K 67/027

(74) Agents: HEFNER, M., Daniel et al.; Leydig, Voit & Mayer, Ltd., Two Prudential Plaza, Suite 4900, 180 North Stetson. Chicago, IL 60601-6780 (US).

(21) International Application Number: PCT/US00/29139

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(22) International Filing Date: 20 October 2000 (20.10.2000)

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(25) Filing Language: English

Published:

— with international search report

(26) Publication Language: English

(88) Date of publication of the international search report:
31 January 2002

(30) Priority Data:

60/161,092 22 October 1999 (22.10.1999) US

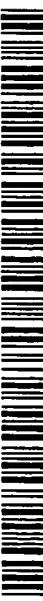
60/227,951 25 August 2000 (25.08.2000) US

(71) Applicant (for all designated States except US): UNIVERSITY OF PITTSBURGH OF THE COMMONWEALTH SYSTEM OF HIGHER EDUCATION [US/US]; 200 Gardner Steel Conference Center, Pittsburgh, PA 15260 (US).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(72) Inventor; and

(75) Inventor/Applicant (for US only): KOIKE, Chihiro [US/US]; 5628 Hempstead Street, Pittsburgh, PA 15206-1520 (US).



WO 01/30992 A3

(54) Title: α 1-3 GALACTOSYLTRANSFERASE GENE AND PROMOTER(57) Abstract: The present invention provides a recombinant expression cassette comprising an α 1-3 galactosyltransferase promoter operably linked to a polynucleotide for expression. The invention also provides a recombinant mutating cassette comprising a region of homology to an α 1-3 galactosyltransferase genomic sequence. The cassettes can be employed to express foreign genes or to disrupt the native α 1-3 galactosyltransferase genomic sequence, particularly within an animal. Thus, the invention also provides transgenic animals and methods for their production and use.

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/29139

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/54 C12N5/10 C12N15/85 C12N9/10 A01K67/027

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N A01K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, BIOSIS, EMBL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 849 991 A (CRAWFORD ROBERT J ET AL) 15 December 1998 (1998-12-15) the whole document ---	13,14, 17-43
X	WO 95 28412 A (GUSTAFSSON KENTH T ;INST OF CHILD HEALTH (GB); BAETSCHER MANFRED W) 26 October 1995 (1995-10-26) claims 1-13 example 6 ---	13,14, 17-43
A	KATAYAMA A. ET AL.: "Porcine a-1,3-galactosyltransferase: full length cDNA cloning, genomic organization and analysis of splicing variants." GLYCOCOCONJ. J., vol. 15, 1998, pages 583-589, XP002159948 the whole document -----	1-43

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

9 February 2001

Date of mailing of the international search report

29.05.2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl
Fax: (+31-70) 340-3016

Authorized officer

Galli, I

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 00/29139

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.: because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

see further information sheet invention 1.

Remark on Protest

- The additional search fees were accompanied by the applicant's protest.
- No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-12 and partly 13-43

A recombinant expression cassette comprising an alpha-1,3-galactosyltransferase promoter operably linked to a polynucleotide for expression, other than a polynucleotide encoding alpha-1,3-galatosyltransferase.

A recombinant mutating cassette comprising a first region of homology to an alpha-1,3-galactosyltransferase genomic sequence adjacent to either a second region of homology to said alpha-1,3-galactosyltransferase genomic sequence or a polynucleotide for insertion, WHEREIN a region of homology is homologous to a promoter of said A alpha-1,3-galactosyl-transferase gene.

Corresponding vectors, recombinant chromosomes, transgenic cells, embryos, organs and animals.

2. Claims: 13-43, partly

As far as not covered by invention 1:

A recombinant mutating cassette comprising a first region of homology to an alpha-1,3-galactosyltransferase genomic sequence adjacent to either a second region of homology to said alpha-1,3-galactosyltransferase genomic sequence or a polynucleotide for insertion.

Said recombinant cassette, wherein the alpha-1,3-galactosyl-transferase sequence is from pig.

Corresponding vectors, recombinant chromosomes, transgenic cells, embryos, organs and animals.

3. Claims: 13-43, partly

Idem as subject matter 2, but wherein the alpha-1,3-galactosyltransferase sequence is from mouse.

4. Claims: 13-43, partly

Idem as subject matter 2, but wherein the alpha-1,3-galactosyltransferase sequence is from man.

INTERNATIONAL SEARCH REPORT

Information on patent family members

In... International Application No

PCT/US 00/29139

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
US 5849991	A 15-12-1998	AU 695373	B	13-08-1998
		BR 9506652	A	02-09-1997
		CA 2181433	A	03-08-1995
		EP 0755451	A	29-01-1997
		AU 1544595	A	15-08-1995
		AU 711144	B	07-10-1999
		AU 7742898	A	01-10-1998
		WO 9520661	A	03-08-1995
		JP 9508277	T	26-08-1997
<hr/>				
WO 9528412	A 26-10-1995	AU 1850599	A	29-04-1999
		AU 2233295	A	10-11-1995
		CA 2187802	A	26-10-1995
		EP 0755402	A	29-01-1997
		JP 10504442	T	06-05-1998
		US 6153428	A	28-11-2000
<hr/>				